Characterisation of a human tRNA gene cluster

David Bourn

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

Department of Biochemistry
University of Glasgow

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

Acknowledgements

Firstly I would like to thank Dr. J.P. Goddard for his supervision of this work and for his consistently helpful and friendly advice over the last three years. Thanks are due to Dr. R.L.P. Adams for useful discussions and for access to his word processor and to Tom Carr for help in the lab. I would also like to thank past and present members of the Department of Biochemistry, and in particular all members of C30 for the Friday evening seminars and for the outrageously long coffee breaks.

I would especially like to thank my parents for their support and encouragement over the years.

Finally, I would like to acknowledge the SERC for financial support.

Abbreviations

The abbreviations used in this thesis are in agreement with the recommendations of the editors of the Biochemical Journal ((1990) *Biochem. J.* **265**, 1-21) with the following additions:

D	dihydrouridine
dNTP	deoxynucleoside-5'-triphosphate
ddNTP	dideoxynucleoside-5'-triphosphate
DMSO	dimethyl sulphoxide
DTT	dithiothreitol
H	hypermodified purine
HEPES	N-2-Hydroxyethylpiperazine-N'-2-ethanesulphonic acid
ICR	internal control region
IPTG	isopropyl-β-D-thio-galactopyranoside
LB	Luria-Bertani
MOPS	3-[N-morpholino] propanesulphonic acid
m.w.	molecular weight
PBS	phosphate buffer saline
pCp	cytidine 3'-5'-biphosphate
PCR	polymerase chain reaction
pfu	plaque forming units
psi	pounds per square inch
rpm	revolutions per minute
tDNA	genes for transfer RNA
TEMED	N,N,N',N',-tetramethylene diamine
TF	transcription factor
VA RNA	virus-associated RNA
X-gal	5-bromo-4-chloro-3-indoyl-β-galactoside
kd	kiloDalton

Contents

Chapter two: Materials and methods.

Chapter five: *In vitro* transcription assays.

Chapter six: Discussion.

List of figures

List of tables

Summary

Prior to the start of this project, several recombinants had been selected from a library of human genomic DNA cloned in λCharon 4A using a mixed tRNA probe (Goddard *et al* ., 1983). One such clone was termed λHt363. Restriction digests of DNA from this recombinant were transferred to nitrocellulose membranes and hybridised to the mixed tRNA probe used in the original cloning. Several fragments from each digest hybridised to the probe, indicating the possible presence of a tRNA gene cluster. The cluster was partially characterised (McLaren and Goddard, 1986) and found to contain a 4.2 kb *Bam* HI fragment carrying three tRNA genes, coding for $tRNA^{Lys}$, $tRNA^{Gln}$ and $tRNA^{Leu}$. The partial sequence of a fourth gene, coding for $tRNA^{Gly}$, was found at one end of an 800 bp *Eco* RI fragment, thought to originate from close to the right hand arm of the λ recombinant.

The restriction map of λHt363 was determined, and the positions of the genes ascertained by correlating the restriction map with the restriction fragments which hybridised to the mixed tRNA probe. The genes and their immediate flanking regions were sequenced, confirming the presence of a $tRNA^{Gly}$ gene. A previously undetected fifth tRNA gene from the recombinant was detected by its transcription *in vitro* and sequenced. This tRNA gene codes for a $tRNA^{Arg}$ with an intron. The presence of an intervening sequence in the $tRNA^{Arg}$ gene possibly explains its poor hybridisation to the mixed tRNA probe. The genes for $tRNA^{Lys}$, $tRNA^{Gln}$ and $tRNA^{Leu}$ have identical coding and almost identical flanking sequences to a cluster previously published (Roy *et al* ., 1982). The gene coding for $tRNA^{Gly}$ has an identical coding and flanking sequence to one published during the course of this study (Doran *et al* . , 1988). The restriction map of λHt363 differs from those of the other published recombinants, suggesting that λHt363 contains a different fragment of genomic DNA. The insert of λHt363 contains a large number of sites for restriction endonucleases with G+C rich recognition sequences. The distribution of these recognition sites, together with sequence data, suggests that the insert contains regions significantly richer in G+C than the norm for human genomic DNA.

Restriction fragments containing tRNA genes, or individual tRNA gene probes, (generated by amplification using the polymerase chain reaction) were employed in Southern hybridisations to human genomic DNA. Such experiments were used to give an estimate of the copy number of the tRNA gene cluster, calculated as 2.7 copies per haploid genome. Probes from the insert of λHt363 hybridised to

genomic restriction fragments which corresponded in size to fragments predicted on the basis of the restriction map of λHt363, suggesting that the insert of the recombinant had not undergone any rearrangements relative to genomic DNA.

Polymerase chain reaction experiments were also used to investigate whether the cloned tRNA gene cluster had undergone any rearrangements, by comparing the sizes of the fragments amplified when using genomic or λHt363 DNA as the target for primers which recognised different sequences within the cluster. A fragment of unexpected size (623 bp) was amplified when using primers specific for the $tRNA^{Gln}$ gene from the tRNA gene cluster on a human genomic DNA target. This fragment was cloned and sequenced and found to contain two human $tRNA^{Gln}$ genes, arranged as a direct repeat. Finally, polymerase chain reaction amplification and direct sequencing of genomic copies of the $tRNA^{Arg}$ gene facilitated the determination of the nature of the intervening sequence in this tRNA gene family.

In vitro transcription assays were used to establish the maximum transcription rate for each of the five tRNA genes from λHt363. The ranking of transcriptional efficiency in this system, from most active to least active gene, was $tRNA^{Gly}$, $tRNA^{Arg}$, $tRNA^{Leu}$, $tRNA^{Gln}$ and $tRNA^{Lys}$. The last four of these tRNA genes were found to have the same relative transcriptional activities when transcribed as a cluster. An investigation into the processing of tRNA precursors in the *in vitro* transcription system, particularly using the $tRNA^{Arg}$ gene, suggested that this system does not efficiently produce mature tRNA species. For the final group of *in vitro* transcription assays, a series of recombinants were constructed with differing amounts of the 5' flank of the $tRNA^{Gly}$ gene deleted. Transcription and competition assays using these recombinants indicated that the 5' flank of this gene may contain an element or elements that affect transcription. Preliminary gel shift and footprinting assays were undertaken to investigate the possibility of any protein or proteins associating with this region of DNA.

CHAPTER 1

INTRODUCTION.

1.1. General introduction.

The major role of transfer RNAs (tRNAs) is to act as adaptor molecules in protein synthesis, although some specific tRNAs have additional functions. They can act as donors of their charged amino acids in the formation of some cell walls and as primers for reverse transcriptase in the synthesis of some retrovirus DNAs. Some tRNAs also have a role in amino acid biosynthesis. The existence of adaptor molecules, needed for the translation of information carried by nucleic acids, was predicted by Crick in 1955. Transfer RNAs were shown to be these adaptors by Hoagland *et al..*, (1958). Recent advances in the techniques of molecular biology have led to a very rapid accumulation of data on the sequences, organisation and expression of tRNA genes. Up to the end of 1988 almost 1000 tRNA genes had been cloned and sequenced (Sprinzl *et al* ., 1989). Results from recent experimental work indicate a high degree of complexity in the fine control of expression and the genomic organisation of these genes. Since tRNAs play a central role in protein synthesis, a better picture of the way in which tRNA gene expression is controlled might be useful in increasing understanding of how organisms meet the demand for the production of proteins, particularly where there are highly tissue- or developmental stage-specific patterns of protein synthesis. The organisation of eukaryotic genomes is another area attracting considerable interest in current research in the field of molecular biology. Studies of the way in which tRNA genes are organised in eukaryotes might help insight into the mechanisms of evolution of complex genomes, and cloned tRNA genes could also be useful as markers for genetic mapping.

1.2. tRNA structure

The nucleotide sequences of some 400 tRNAs and some 1000 tRNA genes have been determined. This primary sequence data provides strong evidence for a secondary structure common to almost all tRNAs, the familiar "cloverleaf" secondary structure shown in figure 1.1. This structure, proposed by Holley *et al..*, (1965) is maintained by base pairing between short complementary regions. Each cloverleaf is made up of a number of domains, consisting of a double helical stem, a single-stranded loop or both a stem and a loop, collectively termed an arm. The amino acid acceptor stem contains the 5' and 3' ends of the tRNA molecule. Usually the stem is made up of seven Watson-Crick base pairs, but a "wobble" G-U pair is present in a yeast $tRNA^{Phe}$ (Clark and Klug, 1975). The 3' terminus consists of four unpaired nucleotides, the last three of which, CCA, are added post-transcriptionally in eukaryotes (see section 1.5.5.). The cognate amino acid is attached to the 3' terminal A. The D arm (see figure 1.1.) is so called because it almost always contains the modified nucleoside dihydrouridine (D). This arm consists of a three or four base pair stem and a variable sized loop. The anticodon arm has a stem of five base pairs and a loop of seven nucleotides. The middle three nucleotides of the loop form the anticodon of the tRNA which pairs with the triplet codon on mRNA during protein synthesis. The variable arm has between four and twenty-one nucleotides, arranged either as a loop of four or five residues or as a stem and loop. Finally, the TΨC arm consists of a five base pair stem and a seven nucleotide loop. The loop always contains the sequence TΨC (Ψ represents pseudouridine). Modified residues are a characteristic feature of tRNAs (Sprinzl *et al.*, 1989).

All tRNAs have invariant nucleotides in the positions indicated in figure 1.1. as well as semi-invariant nucleotide positions (constant purine, (R), or constant pyrimidine, (Y)) (Rich and RajBhandry, 1976). The numbering scheme of the tRNA nucleotides shown in figure 1.1. is based on that of a yeast $tRNA^{Phe}$ which has 76 residues. The total length of tRNAs varies between 74 and 95 nucleotides. However, because the variability in size of tRNAs is confined to three regions of the sequence (two regions in the D loop and the variable arm) a common numbering scheme for all tRNAs is possible.

The three dimensional structure of a yeast $tRNA^{Phe}$ was determined in 1974

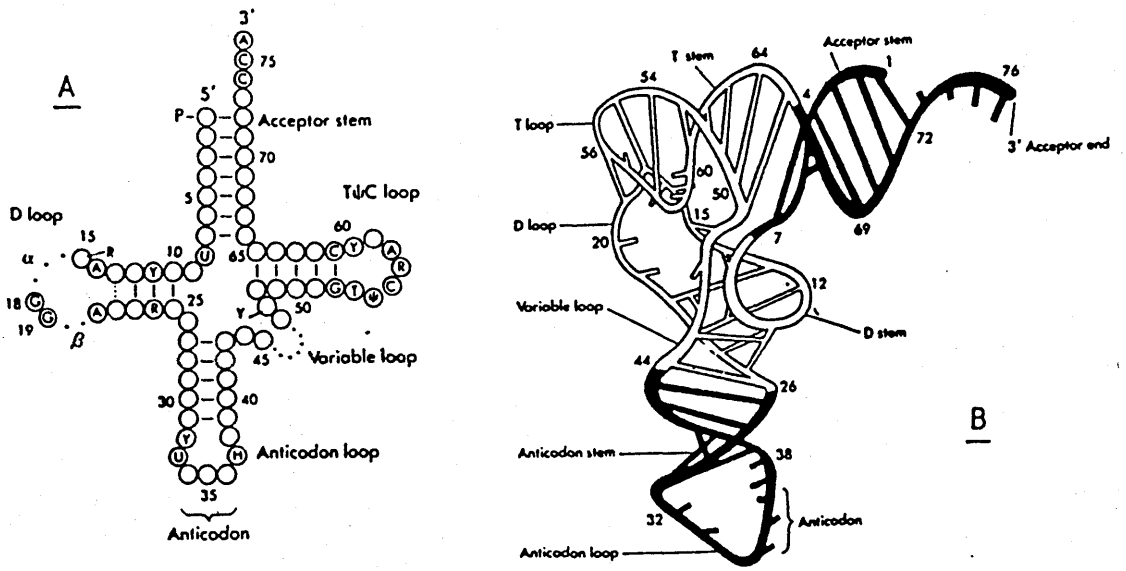
using X-ray diffraction analysis of the crystallised tRNA (Suddath *et al.*, 1974; Robertus *et al.*, 1974). The tertiary structure deduced using this data appears to be common to all tRNAs. Figure 1.1 shows the L-shaped configuration of this structure, formed by tertiary hydrogen bonding interactions, many of which involve invariant or semi-invariant nucleotides. The entire structure is about 2.5 nm thick and 9 nm long (Kim, 1979). The cloverleaf secondary structure is folded into the L shape by the formation of two double helices, one formed by the acceptor stem with the TΨC stem and the other by the D stem with the anticodon stem. The region between the λ coaxial stacked helices contains the TΨC loop and the D loop. A consequence of this conformation is that the amino acid acceptor CCA is located at the far end of the L from the anticodon. This may facilitate alignment of charged tRNAs on the ribosome, although this has not been proved.

1.3. tRNA identity.

In order for protein synthesis to occur correctly, each tRNA must be aminoacylated with the appropriate amino acid through the action of an aminoacyl-tRNA synthetase. Because of degeneracy in the genetic code, most amino acids are represented by a group of tRNAs, termed isoacceptors. In general there is a single aminoacyl-tRNA synthetase for each amino acid. Each group of isoaccepting tRNAs must share a set of features enabling recognition only by the cognate aminoacyl-tRNA synthetase. These distinguishing elements comprise the identity of the tRNA. Recently, two complementary approaches have been used in determining the mechanisms of tRNA identity, one involving the alteration of tRNA suppressor identity and the other the production of altered tRNAs *in vitro* using transcription of synthetic tRNA genes (reviewed in Normanly and Abelson, 1989). Such studies have produced two main conclusions. Firstly, it appears that tRNA identity is determined by a relatively small number of elements, and these can act either positively in determining the identity of a tRNA or negatively in preventing recognition by an aminoacyl-tRNA synthetase. Secondly, the anticodon appears to be an important identity element for a majority of tRNAs. The anticodon would seem to be the most logical identity element since it is directly responsible for association with the appropriate codon. However, it has long been established that the anticodon cannot control the identity for all tRNAs because of the existence of amber

Figure 1.1. Secondary and tertiary structure of tRNA.

Figure 1.1.A shows the general cloverleaf secondary structure for tRNAs with the standard numbering system based on that for yeast $tRNA^{Phe}$. The positions of invariant and semi-invariant bases are shown. Y denotes pyrimidine, R purine, Ψ pseudouridine and H hypermodified purine. The dotted regions in the D loop and in the variable loop contain different numbers of residues in different tRNAs.
(Adapted from Watson *et al* ., 1987).
Figure 1.1.B illustrates the tertiary structure of yeast $tRNA^{Phe}$. The ribose-phosphate backbone is drawn as a continuous ribbon and internal hydrogen bonding is indicated by crossbars. The positions of unpaired bases are indicated by unjoined bars. The anticodon arm and the acceptor stem are shaded.
(Adapted from Watson *et al* ., (1987), after Kim *et al* ., (1974)).



suppressor tRNAs where the anticodon has been altered with no effect on amino acid specificity. Schulman and Pelka (1990) used synthetic tRNA genes to show that the anticodon is a major recognition element in *E. coli* $tRNA^{Met}$. The anticodon is also an important element for *E. coli* $tRNA^{Val}$ identity (Schulman and Pelka, 1988).

The number of residues responsible for the identity of a tRNA (the "identity set") appears to differ for different tRNA species. Hou and Schimmel (1988) showed that in *E. coli* $tRNA^{Ala}$, mutations in the G-U pair at position 3-70 (found in almost all $tRNA^{Ala}$ species sequenced) abolished recognition by the alanine aminoacyl-tRNA synthetase. Insertion of G_3-U_{70} was sufficient to cause a $tRNA^{Cys}$ amber suppressor to insert alanine. In contrast, the identity set for *E. coli* $tRNA^{Ser}$ appears to involve at least 8 residues (Normanly *et al* ., 1986, Normanly and Abelson, 1989, Himeno *et al* ., 1990) and that for yeast $tRNA^{Phe}$ at least 5, of which three comprise the anticodon (Sampson *et al* ., 1989). As yet, no general rules have been established for identity determination.

1.4. Eukaryotic tRNA gene organisation.

Eukaryotic transfer RNA genes are members of complex multigene families. The tRNA genes are found dispersed throughout the eukaryote genome, either as solitary genes or in small clusters. There may be several different isoaccepting tRNAs for each amino acid, which implies that there is a similar redundancy in tRNA gene species. The gene for any particular tRNA isoacceptor may be present in multiple copies in the genome, and these gene copies may be located at more than one chromosomal locus. It is uncertain how different gene copies within gene families originated or how they are maintained, but it is probable that some new gene copies will have been produced by gene duplication. There is some evidence that tRNA genes can be inserted at new sites in the genome as a result of retroposition events (Reilly *et al* ., 1982; McBride *et al* ., 1989). Gene copies may be maintained by gene conversion since it is difficult to see how selective pressures could act on individual members of a gene family containing ten or more genes. Several hundred eukaryotic tRNA genes have been cloned, either as dispersed individual genes or in small

clusters of two or more genes within a few kilobases of each other. The tRNA genes within these clusters may code for different tRNAs and can occur on either DNA strand. Unlike *E. coli* tRNA genes, these clustered genes are generally not cotranscribed as large multimeric precursors although there are examples of pairs of eukaryotic tRNA genes being transcribed as dimeric precursors (Schmidt *et al* ., 1980; Mao *et al* ., 1980; Willis *et al* ., 1984; Amsutz *et al* ., 1985).

So far, tRNA genes from 17 different eukaryote species have been sequenced. Most data is available for the organisms described in the following sections: yeast, *Drosophila melanogaster*, *Xenopus laevis*, rat, mouse and finally humans.

1.4.1. Organisation of yeast tRNA genes.

Saturation hybridisation analysis of *S. cerevisiae* genomic DNA gives an estimate of 360 tRNA genes per haploid genome, which corresponds to an average copy number of 8 per tRNA species (Schwiezer *et al* ., 1969). Members of at least ten yeast tRNA gene families contain introns (Ogden *et al* ., 1984; Stuka and Feldmann, 1988). Yeast tRNA genes are frequently associated with transposable elements such as *sigma*, *delta*, Ty1 and *tau* (Eigel and Feldmann, 1982; Hauber *et al* ., 1988).The most common of these is the *sigma* element, which has been found to occur 16 to 18 bp upstream of several yeast tRNA genes (Sandmeyer and Olsen, 1982; Del Ray *et al* ., 1982; Brodeur *et al* ., 1983; Sandmeyer *et al* ., 1988). *Sigma* is about 340 bp long, with 8 bp inverted terminal repeats and 5 bp flanking direct repeats. *Sigma* elements have so far only been found in association with tRNA genes, which has led rise to speculation that they may play a role in the evolution or the transcriptional initiation of yeast tRNA genes.

Most yeast tRNA genes appear to be dispersed randomly through the genome although some clusters have been cloned. For example, a $tRNA^{Ser}$-$tRNA^{Met}$ gene pair has been found at three different loci in *S. pombe* . The genes are separated by 7 bp and are transcribed as a dimeric precursor, which is unusual for eukaryotic tRNA genes (Mao *et al* ., 1980; Amsutz *et al* ., 1985). Similarly, four separate clones have been characterised in which genes for $tRNA^{Arg}$ and $tRNA^{Asp}$ are separated by 9 bp and are again transcribed as a dimeric precursor (Schmidt *et al* ., 1980).

Probably the best characterised dispersed yeast tRNA gene family is that for a *S.*

cerevisiae $tRNA^{Glu}$ gene (Hauber *et al* .,1988). Eight different gene copies have been cloned. These copies all have the same coding sequence, contain no introns and are flanked by transposable elements. The genes have been assigned to different chromosomal loci and show little or no homology in their flanking sequences. The eight $tRNA^{Tyr}$ genes from *S. cerevisiae* again have identical coding sequences (although there is a C/T polymorphism within the intron) but differ in their flanking regions (Johnson and Abelson, 1983).

1.4.2 Organisation of *Drosophila melanogaster* tRNA genes.

The *D. melanogaster* genome contains some 750 tRNA genes, (Weber and Berger, 1976) with an average copy number of 12 for each of the approximately 60 isoacceptor species. *In situ* hybridisation studies using polytene chromosomes (Gall and Pardue, 1969; Steffensen and Wimber, 1971) have shown these genes to be distributed over at least 50 sites in the genome, with clustering at particular loci. Clusters may be of multiple copies of the same gene (homoclusters) or of different genes (heteroclusters).

Chromosomal locus 42A of the *Drosophila* chromosome 2R contains a region of 94kb isolated by Yen and Davison (1980). Within this 94 kb area, a central 46kb region contains three widely spaced homoclusters of $tRNA^{Asn}$, $tRNA^{Arg}$ and $tRNA^{Lys}$ genes. Another homocluster characterised consists of five nearly identical copies of $tRNA^{Glu}_{CTC}$ (Hosbach *et al* ., 1980). These genes are organised as a group of three genes spanning approximately 0.55 kb, followed by a 0.45 kb gene doublet 3 kb downstream. The flanking regions of certain of the gene copies showed strong patterns of sequence homology. The pattern of homology led Hosbach *et al* . to propose that this $tRNA^{Glu}$ homocluster evolved through two ancestral $tRNA^{Glu}_{CTC}$ genes each giving rise to gene doublets by gene duplication, and one of these gene pairs then forming a trio of genes as the result of an unequal crossover. A third homocluster at chromosomal locus 12DE consists of eight $tRNA^{Ser}$ genes (for $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$) . The genes are 96% homologous, and all of the gene copies show several blocks of homology in their flanking sequences. A ninth $tRNA^{Ser}$ gene, separate from the gene cluster has also been sequenced (Cribbs *et al* ., 1987). Of the six *Drosophila* $tRNA^{Tyr}$ genes

characterized, three compose a homocluster and two a second homocluster. The sixth gene copy is a solitary gene. All six genes have identical coding sequences but contain different sizes of introns, and no apparent homology exists between the flanks of the different gene copies (Suter and Kubli, 1988).

In addition to the solitary $tRNA^{Ser}$ and $tRNA^{Tyr}$ genes mentioned above, several other tRNA genes from *Drosophila* do not appear to be organised in clusters. Genes for $tRNA_i^{Met}$ at chromosomal locus 61D are separated by at least 17 kb (Sharp *et al* ., 1981a), and solitary $tRNA^{Gly}$ genes have been mapped to chromosomal locus 56F (Hershey and Davison, 1980). The irregular dispersed arrangement of tRNA genes in *Drosophila* has turned out to be the general case for higher eukaryotes.

1.4.3. Organisation of *Xenopus laevis* tRNA genes.

Generally, the copy number of tRNA genes varies from organism to organism, and is only approximately related to genome size. The *Xenopus laevis* genome contains a very large population of tRNA genes (6500-7800 genes per haploid complement; see Long and Dawid, 1980). The majority of gene copies are located within DNA fragments repeated 150 to 300-fold (Clarkson and Kurer, 1976). One of these repeated DNA fragments contains a cluster of eight tRNA genes within a 3.18 kb region of DNA which is tandemly repeated approximately 150 times per haploid genome (Clarkson *et al* ., 1978; Muller and Clarkson, 1980; Fostel *et al* ., 1984; Muller *et al* ., 1987). The tandem repeats probably form a large gene cluster on one chromosome. Each repeating unit contains two $tRNA^{Met}$ genes, the A-copy and the B-copy, along with one gene each for $tRNA^{Phe}$, $tRNA^{Tyr}$, $tRNA^{Asp}$, $tRNA^{Ala}$, $tRNA^{Leu}$ and $tRNA^{Lys}$. The tRNA coding regions are separated by large regions of noncoding, spacer DNA. Both the coding and noncoding regions show sequence homology between repeating units. Another highly reiterated gene, for $tRNA^{Val}$, has been found on two 250-fold repeated fragments, of 892 and 4100 bp (Peterson, 1987). However, a dispersed $tRNA^{Tyr}$ gene, not highly repeated in the *Xenopus* genome, has also been characterised. The gene differs from the clustered $tRNA^{Tyr}$ gene mentioned above in both its coding and flanking regions (Gouilloud and Clarkson, 1986).

1.4.4. Organization of the rat and mouse tRNA genes.

Other than in humans, mammalian tRNA gene organisation has been studied mainly in the rat and the mouse. In both species tRNA genes are found as solitary gene copies as well as in small clusters. In the rat, the average copy number of tRNA genes is 10-20 per haploid genome (Lasser-Wiess *et al* ., 1981). In the mouse, the average copy number may be as high as 100 (Marzluff *et al* ., 1975). Most rat tRNA genes characterized so far occur in small clusters, although this may be a function of the relative ease of detection of a tRNA gene cluster as opposed to a single tRNA gene. One such cluster was isolated on a 3.3 kb *Eco* RI fragment.The cluster contains genes for $tRNA^{Leu}$, $tRNA^{Asp}$, $tRNA^{Gly}$ and $tRNA^{Glu}$, and appears to be part of a larger 13.5 kb repeated unit, with a copy number of 10 per haploid genome (Sekiya *et al* ., 1981; Shibuya *et al* ., 1982 and 1985). When the sequences for all 10 repeats were compared, seven appeared to be broadly homologous, although deletions, insertions or base substitutions were observed at several positions (Rosen *et al* ., 1984). Five of the repeats contain tRNA pseudogenes (Shibuya *et al* ., 1982). One copy of the cluster (containing a $tRNA^{Leu}$ pseudogene instead of the $tRNA^{Leu}$ gene) showed only minimal homology to the rest of the repeats. This data led Rosen *et al* ., (1984) to suggest that the different copies of the cluster were derived by gene duplication, followed by mutations and deletions of nonessential sequences. Interestingly, a mouse tRNA gene cluster has been characterised which is highly homologous to one of the repeats in the rat. The rat and mouse $tRNA^{Asp}$ and $tRNA^{Gly}$ genes are identical, while the mouse $tRNA^{Glu}$ gene differs from that on the rat cluster in only two residues (Sekiya *et al* ., 1981; Looney and Harding, 1983). A solitary $tRNA^{Asp}$ gene, identical to that in the rat cluster, has also been characterised from the mouse (Looney and Harding, 1983).

Several other small tRNA gene clusters have been found in rat and mouse. One cluster in the rat contains three $tRNA^{Lys}$ genes and three $tRNA^{Pro}$ genes. The genes occur on both strands and show no homology in their 5' flanking sequences (Sekiya *et al* ., 1982). In the mouse, a partially characterised cluster has been shown to carry a $tRNA^{Pro}$ and a $tRNA^{Gly}$ gene (Hu *et al* ., 1983) and a cluster carrying single copies of genes for $tRNA^{Ala}$, $tRNA^{Ile}$, $tRNA^{Pro}$ and $tRNA^{Lys}$ has been reported (Russo *et al* ., 1987a).

In addition to the solitary mouse $tRNA^{Asp}$ gene mentioned above, scattered tRNA genes have been reported for a mouse $tRNA^{His}$ gene (Han and Harding, 1982); a mouse $tRNA_i^{Met}$ gene (Han *et al* ., 1984) and a rat $tRNA^{Phe}$ gene (Rosen and Daniel, 1988).

1.4.5. Organisation of human tRNA genes.

The human haploid genome contains in the order of 1000 tRNA genes (Hatlen and Attardi, 1971), of some 60 different species (Lin and Agris, 1980). This implies an average copy number for each individual gene of 10-20. Copy numbers have been determined for several human tRNA gene families using Southern hybridisations of the genes to restriction fragments from genomic DNA. Most of the published estimates for copy number fall within the predicted range. Among these estimates are 13 copies per haploid genome for $tRNA^{Glu}$ (Gonos and Goddard, 1990a), 12 copies per haploid genome for $tRNA_i^{Met}$ (Zasloff and Santos, 1980), 13-16 copies for $tRNA^{Lys}$ (Doering *et al* ., 1982), at least 8 copies for $tRNA^{Ser}$ (Krupp *et al* ., 1988) and at least 13-15 copies for the human $tRNA^{Val}$ gene family (Arnold *et al* ., 1986; Schmutzler and Gross, 1990). However, the copy number for a human $tRNA^{Asn}$ gene family was estimated as 60-70 per haploid genome (Buckland, 1989). This level of reiteration has not been found for any other human tRNA gene family.

As in other mammalian systems, human tRNA genes have been isolated both as dispersed gene copies and as members of small gene clusters of 2-5 genes within a few kb of each other. Over 50 human tRNA genes have been cloned (Schmidtke and Cooper, 1990), and a number of these have been assigned to their respective chromosomes. Assignments have been made either by *in situ* hybridisations or using somatic cell hybrids with differing complements of human chromosomes. Transfer RNA genes have been assigned to 8 different chromosomes so far, and given the large number of human tRNA genes and the lack of any apparent pattern of organisation, it seems likely that most if not all human chromosomes will be found to carry tRNA genes. Several tRNA gene clusters have been mapped to single sites in the genome (see below). In contrast, two $tRNA^{Gly}_{GCC}$ genes with an identical coding sequence but different flanks have been assigned to two different

chromosomes (see below and Chapter four).

The human $tRNA_i^{Met}$ gene family has been extensivly studied (Zasloff and Santos, 1980; Santos and Zasloff, 1981; Zasloff *et al* ., 1982; Wahab *et al* ., 1989). Twelve genes have been found at scattered locations through the genome. Analysis of two of these gene copies showed the genes to be identical, with several blocks of homology in the flanking regions. Both are located on chromosome 6 (6p23-q12) (Naylor *et al* ., 1983).

$tRNA^{Tyr}$ genes have also been detected at at least 12 independent loci in the human genome (McPherson and Roy, 1986; van Tol *et al* ., 1987). All of these genes have been shown to contain an intron (van Tol and Beier, 1988). The four gene copies characterised have almost identical coding sequences. Two of these characterised $tRNA^{Tyr}$ gene copies belong to homologous 200 bp repeat fragments (McPherson and Roy, 1986).

Thirteen independent $tRNA^{Val}$ gene loci have been detected in the human genome. Studies on three of the members of the $tRNA^{Val}$ gene family have shown two to encode the major and minor $tRNA^{Val}$ isoacceptors, while the third does not correspond to any known $tRNA^{Val}$. The flanks of these genes do not show any significant homology. Two of the gene copies are associated with *Alu* type repeats (Arnold *et al* ., 1986). Six additional human $tRNA^{Val}$ genes have been cloned and characterised (Thomann *et al* .,1989).

Three different human genomic clones have been isolated with dispersed $tRNA^{Gly}_{GCC}$ genes (in addition to a fourth human $tRNA^{Gly}_{GCC}$ gene described in Chapter three of this study). All three have the same coding sequence but different flanking sequences (Pirtle *et al* ., 1986; Doran *et al* ., 1988). The $tRNA^{Gly}_{GCC}$ gene cloned by Pirtle *et al* . is 758 bp away from a $tRNA^{Gly}_{GCC}$ pseudogene, and the DNA fragment carrying the $tRNA^{Gly}_{GCC}$ gene and pseudogene has been mapped to chromosome 16 (McBride *et al* ., 1989). A solitary $tRNA^{Gly}_{CCC}$ gene has also been cloned, (Shortridge *et al* ., 1985) and assigned to chromosome 1 (1p34-pter) (McBride *et al* ., 1989).

A solitary human $tRNA^{Glu}$ gene has been cloned and mapped to chromosome 1 (1p36 and a secondary site at 1q21) (Goddard *et al* ., 1983; Boyd *et al* ., 1989). A second solitary $tRNA^{Glu}$ gene, with a coding sequence differing in two non-conserved residues from the first and similar 5' and 3' flanks has also been characterised (Gonos and Goddard, 1990b). Among the other dispersed human tRNA genes cloned are two $tRNA^{Asn}$ genes (Ma *et al* ., 1984), four $tRNA^{Ser}$ genes

(Yoo, 1984; Hong *et al* ., 1987; Hoe *et al* ., 1987; Krupp *et al* ., 1988) and finally an opal suppressor phosphoserine tRNA gene. This last gene has been assigned to chromosome 19, and its corresponding pseudogene to chromosome 22 (O'Neil *et al* ., 1985; McBride *et al* ., 1987).

A number of small tRNA gene clusters have been cloned from human genomic DNA. One such cluster contains three tRNA genes, coding for $tRNA^{Lys}$, $tRNA^{Gln}$ and $tRNA^{Leu}$, separated by short (about 500 bp) intergenic regions (Roy *et al* ., 1982; Buckland *et al* ., 1983). (See Chapter 3 for a description of an almost identical copy of this cluster). A second cluster characterised contains four tRNA genes arranged in two pairs, separated by 3 kb. A $tRNA^{Pro}$ gene is separated from a $tRNA^{Leu}$ gene by 724 bp in the first pair and a second $tRNA^{Pro}$ is 316 bp away from a $tRNA^{Thr}$ gene in the second pair (Chang *et al* ., 1986). This cluster has been mapped to chromosome 14 (Olsen *et al* ., 1987). The $tRNA^{Pro}$ genes have identical coding sequences but differ in their flanking regions. Another cluster of four tRNA genes contains two genes for $tRNA^{Lys}$ and two for $tRNA^{Phe}$ arranged in alternating order. The $tRNA^{Lys}$ genes are identical, and the $tRNA^{Phe}$ genes differ in only one nucleotide, at residue 57. Again, no apparent homology exists in the flanking regions of the genes (Doran *et al* ., 1987). Two further clusters containing three human tRNA genes have been characterised. One encodes the major and minor valine tRNAs and a lysine tRNA. The recombinant carrying these genes also contains at least 9 *Alu* family members (Craig *et al* ., 1989). The final cluster encodes threonine, proline and valine tRNAs. This cluster is associated with two *Alu* elements, one of which overlaps the $tRNA^{Val}$ gene (Shortridge *et al* ., 1989). This cluster has been mapped to chromosome 5 (McBride *et al* ., 1989).

1.4.6. tRNA pseudogenes.

Given the large number of tRNA genes occurring in eukaryotic genomes, it is perhaps not surprising that a considerable number of tRNA pseudogenes have been characterised. For a protein-producing gene, a pseudogene can be defined as one that gives rise either to no product or to an inactive polypeptide. For tRNA genes, the term generally implies a sequence which if transcribed cannot be folded into a cloverleaf configuration, and that does not correspond to any known RNA species.

tRNA pseudogenes may consist either of a partial sequence of a corresponding tRNA sequence or of sequences that resemble a tDNA with differences at some positions. Transfer RNA genes with mutations in one or more of the promoter elements, rendering them transcriptionally inactive, might also be classed as pseudogenes. Pseudogenes have been sequenced in several eukaryotes, although so far none are known in yeast. The significance, if any, of this finding is uncertain.

In *Drosophila melanogaster*, several pseudogenes have been characterised. A $tRNA^{His}$ gene is clustered with a $tRNA^{His}$ pseudogene. The pseudogene contains a run of 8 nucleotides between residues 38 and 45 that differ from the corresponding gene sequence. However, this gene was found to be transcriptionally active *in vitro* when its 5' flanking sequence was replaced with one from an active $tRNA^{His}$ gene (Cooley *et al* ., 1984). (Other eukaryotic pseudogenes have been found to be transcriptionally inactive *in vitro* ; Shibuya *et al* ., 1982; Sharp *et al* ., 1981a).

A dispersed "family" of pseudogenes has been located at about 30 sites in the *Drosophila* genome. These sites all contain short fragments which are homologous to parts of the initiator tRNA. The largest of these only spans 34 bp, representing approximately half of the intact sequence. The fragments have a genomic distribution characteristic of a mobile DNA element, suggesting that this "family" may have arisen through the repeated insertion and excision of a transposable element into an intact tRNA gene (Sharp *et al* .,1981a).

A number of pseudogenes have been found within a rat tRNA gene cluster (Sekiya *et al* ., 1981; Shibuya *et al* ., 1982 and 1985; see section 1.4.4.). In the ten characterised repeats of the cluster (encoding $tRNA^{Leu}$, $tRNA^{Asp}$, $tRNA^{Gly}$ and $tRNA^{Glu}$) a total of 9 pseudogenes have been discovered, including incomplete $tRNA^{Gly}$ and $tRNA^{Glu}$ genes and a $tRNA^{Leu}$ pseudogene which differs in 15 residues from the corresponding $tRNA^{Leu}$ gene (Rosen *et al* ., 1984; Shibuya *et al* ., 1982; Makowski *et al* ., 1983).

A $tRNA^{Phe}$ pseudogene has been isolated from the mouse (Reilly *et al* ., 1982). This pseudogene consists of the 38 nucleotides of the 3' end of the tRNA, including the terminal CCA. The terminal CCA is added post-transcriptionally to all eukaryotic tRNAs for which the genes have been identified. This suggests that the psuedogene may have arisen through a tRNA gene product being partially copied back into the genome. A mouse $tRNA^{Glu}$ pseudogene has also been characterised (Smardo and Calvert, 1987).

Two human tRNA pseudogenes were mentioned in section 1.4.5. One of these, a

human opal suppressor phosphoserine tRNA pseudogene differs in 7 nucleotides from the corresponding "true" gene (O'Neil *et al* ., 1985). The other, a $tRNA^{Gly}_{GCC}$ pseudogene differs in just 2 nucleotides (including the invariant C at position 56) from the corresponding gene (Pirtle *et al* ., 1986). The two $tRNA^{Gly}$ genes share 8 nucleotides of 5' flanking sequence.

1.5. Transcription of eukaryotic tRNA genes.

Both *in vivo* and *in vitro* experimental systems have been developed to study tRNA gene transcription. Transcription assays can be performed using micro-injection of DNA into *Xenopus* oocyte nuclei or using cell-free transcription systems. Such studies have been used to show that the primary product of tRNA gene transcription is not the tRNA itself but a longer precursor molecule. This precursor is processed in a stepwise manner to yield the mature tRNA (see section 1.5.5.). Transcriptional studies on truncated or mutated tRNA genes have allowed the characterisation of the tRNA gene promoter and the initiation and termination sites for transcription. These studies, coupled with experiments involving the use of purified transcription factors, have also helped to elucidate the mechanism by which the tRNA gene interacts with transcription factors to form the transcription complex.

1.5.1. Genes transcribed by RNA polymerase III.

Eukaryotic RNA polymerase III is a multi-subunit enzyme of approximately 650 kd (for review see Ciliberto *et al* ., 1983a; Geiduschek and Tocchini-Valentini, 1988). RNA polymerase III synthesizes a variety of small nuclear and cytoplasmic RNAs. Among these are tRNA, 5S ribosomal RNA, 7S cytoplasmic RNA and U6 and 4.5S small nuclear RNAs. The enzyme also transcribes a number of genes coding for small RNAs in adenoviruses (VAI and VAII), and in Epstein-Barr virus (EBER I and EBER II). In addition RNA polymerase III is responsible for the synthesis of small RNAs from some repetitive elements (e.g. the *Alu* family).

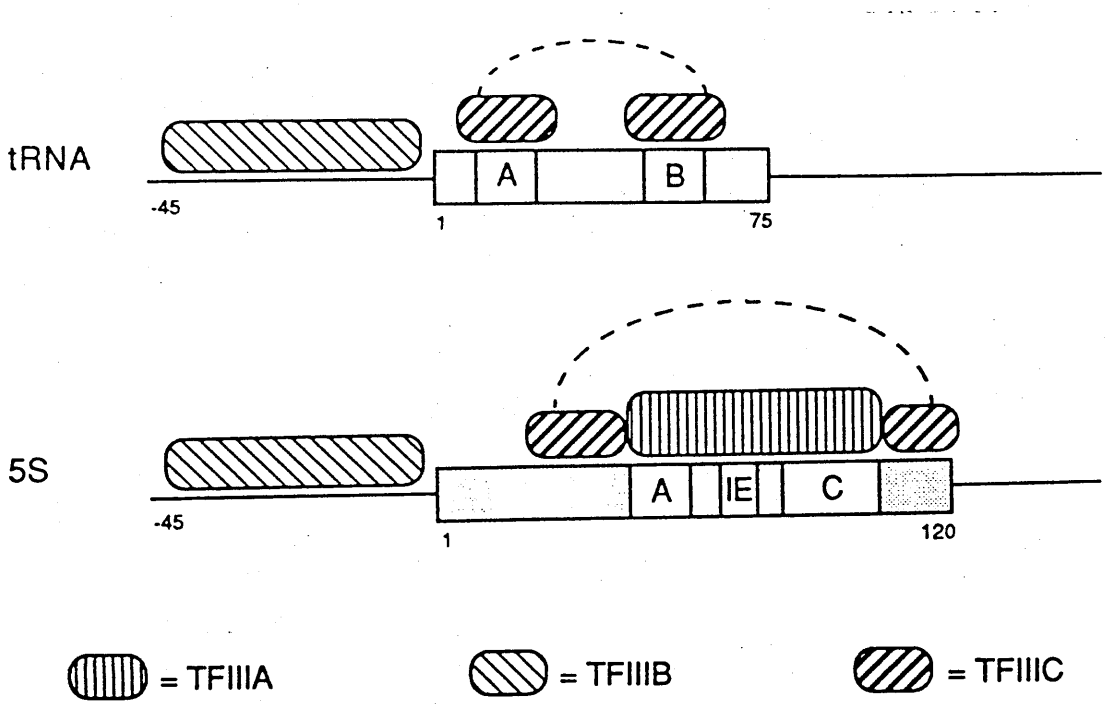
Transcription of the majority of RNA polymerase III genes is controlled by intragenic DNA sequences (Sharp *et al* ., 1985; Geiduschek and Toccini-Valentini,

1988 and references therein). These internal control regions (ICR) are discontinuous, with conserved "boxes" separated by spacer regions. RNA polymerase III genes can be divided into two main groups according to the nature of these internal promoter sequences. Type 1 genes (5S RNA) have an ICR consisting of 2 domains, box A and box C, with an intermediate element associated with box C. Type 2 genes (e.g. tRNA genes) again have two domains in their internal promoters, box A and box B. The box A sequence is highly conserved in both types of promoter, and the box B sequence is highly conserved in type 2 promoters. The conservation of promoter sequences between different RNA polymerase III genes explains the fact that only three major transcription factors (TFIIIA, TFIIIB and TFIIIC) are required for the transcription of all of the genes transcribed by this enzyme.The promoters of VA and *Alu* family small RNA genes belong to type 2, and sequence homologies suggest that the EBER genes, human 7SL and mouse 4.5S genes also belong to this class. The U6 and 7SK genes have promoter sequences 5' to the gene, and therefore fall into a separate group (for review see Folk, 1988; Sollner-Webb, 1988). The type 2 promoter of tRNA genes will be discussed in section 1.5.2.

The intragenic 5S promoter has been extensively studied. Genetic dissection of a *Xenopus* somatic 5S RNA gene established that its promoter extends from residues 50-97 of the gene, this region containing 3 separate elements that determine promoter strength. (Sakonju *et al* ., 1980; Bogenhagen *et al* ., 1980; Pieler *et al* ., 1985a, 1985b and 1987). At least three transcription factors (TFIIIA, TFIIIB and TFIIIC; Segall *et al* ., 1980; Lassar *et al* ., 1983; Bieker *et al* ., 1985; Setzer and Brown, 1985) are required to bind to the promoter elements prior to the addition of RNA polymerase III. The box A element (residues 50-64) is involved in the binding of TFIIIC, while the intermediate element (residues 67-72) and box C (residues 80-97) have a high affinity for TFIIIA. The binding of TFIIIA (which contains zinc finger motifs) to the gene helps recruit TFIIIB and TFIIIC to the stable transcription complex (Engelke *et al* ., 1980; Sakonju *et al* ., 1981; Bogenhagen *et al* ., 1982; Brown, 1984; Schlissel and Brown, 1984). (TFIIIA has no effect on transcription of type two genes ; Gottesfeld and Bloomer, 1982). The sequence of the spacer regions between the elements is unimportant but alterations in their length can reduce or prevent the formation of a stable transcription complex (Pieler *et al* ., 1987; Bogenhagen, 1985). Figure 1.2. summarises the transcription factor binding sites found in 5S RNA genes.

Figure 1.2. Transcription factor binding sites on tRNA and 5S RNA genes.

The sites on tRNA and 5S RNA genes that are protected from DNase I digestion by the RNA polymerase III transcription factors TFIIIA, TFIIIB and TFIIIC are indicated below. Boxes A, B and C and the intermediate element (IE) are shown. Dotted lines indicate the possible existence of two different subunits of a single large factor. (From Palmer and Folk, 1990).



1.5.2. The tRNA gene promoter.

Analyses of the transcription of a number of eukaryotic tRNA genes, including a *Xenopus* $tRNA^{Met}$ gene (Kressmann *et al* ., 1979; Hofstetter *et al* ., 1981; Folk and Hofstetter, 1983) and $tRNA^{Leu}$ gene (Galli *et al* ., 1981), a *Drosophila* $tRNA^{Arg}$ gene (Sharp *et al* ., 1981b, 1982, 1983a and 1983b) and $tRNA^{Lys}$ gene (De Franco *et al* ., 1980), a *C. elegans* $tRNA^{Pro}$ gene (Traboni *et al* ., 1984) and a yeast suppressor $tRNA^{Tyr}$ gene (Koski *et al* ., 1980 and 1982; Allison *et al* ., 1983), have defined an internal promoter with two domains. Internal deletions and substitutions in these genes have located the promoter domains, box A and box B, to residues 8-19 and 52-62 respectively (numbered according to the standard system described in section 1.2.). The box A sequence in tRNA and other type two gene promoters is structurally and functionally homologous with the box A component of the 5S gene promoter (Ciliberto *et al* ., 1983b; Geiduschek and Tocchini-Valentini, 1988). Box A is also termed the D control region or 5' ICR, and box B the T control region or 3' ICR. The two regions of the ICR coincide with the highly conserved nucleotides in the D and TΨC arms of the tRNA gene. Genetic manipulations of these residues in conjunction with transcription assays have been used to determine the consensus sequence of the promoter: TRGCNNAGY-GG for box A and GGTTCGANTCC for box B (where N= any nucleotide, R= a purine, Y= a pyrimidine and - a nucleotide (residue 17) not present in all tRNAs; Geiduschek and Tocchini-Valentini, 1988). The box B sequence is the more highly conserved of the two, box A showing some variation both in sequence and in the number of nucleotides. This high degree of conservation of box B in all characterised tRNA genes may reflect constraints due to the role of these residues in the tertiary structure of the tRNA product as well as in interactions with transcription factors.The spacing between box A and box B varies in tRNA genes owing to the presence of intervening sequences and the varying lengths of the extra arm in some genes. In yeast genes the spacing varies between 31 and 93 bp (Raymond and Johnson, 1983).

1.5.3. Transcription complex formation.

It has been shown that tRNA genes are efficient competitors of adenovirus VA I genes as well as of type one genes (5S RNA) which suggests that faithful transcription of these genes depends on similar transcription factors (Segall *et al* ., 1980; Burke *et al* ., 1980; Wormington *et al* ., 1981; Guilfoyle and Wienmann, 1981). Transcription studies have established that eukaryotic tRNA genes are transcribed by RNA polymerase III in the presence of at least two transcription factors, TFIIIB and TFIIIC (Stillman and Geiduschek, 1984; Baker and Hall, 1984; Baker *et al* ., 1986; Huibregtse *et al* .,1987; Burke and Soll, 1985; Carey *et al* ., 1986). (In yeast, transcription factor τ is analogous to the vertebrate factor TFIIIC). TFIIIB has been purified from *S. cerevisiae* as a 60 kd polypeptide (Klekamp and Weil, 1986a and 1987). Antibodies to this polypeptide have been shown to inhibit transcription. Since TFIIIB alone cannot bind stably or specifically to DNA (Lassar *et al* ., 1983; Klekamp and Weil, 1986b; Carey *et al* ., 1986), its incorporation into transcription complexes is dependent on interactions with other proteins.

TFIIIC binds tightly and specifically to DNA, thereby sequestering and directing transcription activity (Lasser *et al* ., 1983). Vertebrate TFIIIC has been only partially purified (from HeLa cells, Schneider *et al* ., 1989), but the corresponding yeast factor τ has been extensively characterized (Ruet *et al* ., 1984; Camier *et al* ., 1985; Kassevetis *et al* ., 1989; Gabrielsen *et al* ., 1989a and 1989b; Johnson and Wilson, 1989; Bartholemew *et al* ., 1990), and shown to be a multisubunit protein of about 600 kd. TFIIIC and τ have different molecular properties, although they seem to perform the same function. TFIIIC from HeLa cells can be resolved by chromatography into two components, TFIIIC1 and TFIIIC2 (Yoshinaga *et al* ., 1987; van Dyke and Roeder, 1987; Dean and Berk, 1987). These may represent separate factors which copurify. In contrast, yeast factor τ cannot be resolved in the same way into two active components. It has been suggested that this protein contains two domains, τ_A and τ_B (Gabrielsen *et al* ., 1989b; Marzouki *et al* ., 1986) which may be analogous to TFIIIC1 and TFIIIC2. However, photochemical crosslinking experiments (Bartholomew *et al* ., 1990), have shown that TFIIIC appears to be made up of subunits of 145, 135, 95 and 55 kd. The 135 kd subunit is thought to play a role in positioning the one of the subunits of TFIIIB on the transcription complex.

The physical interactions of TFIIIC/τ have been analysed qualitatively by various footprinting methods (Camier *et al* ., 1985; Stillman *et al* ., 1985; Klemenz *et al* ., 1982; Fuhrman *et al* ., 1984; Ruet *et al* ., 1984; Stillman and Geiduschek, 1984) and quantitatively by gel retardation assays (Baker *et al* ., 1986 and 1987). DNA protection experiments have shown TFIIIC2 to bind to the box B region of the tRNA gene promoter. The addition of TFIIIC1 to the TFIIIC2-DNA complex extends DNaseI protection to the box A region (Yoshinaga *et al* ., 1987), and the binding of TFIIIB creates a footprint that includes approximately 40 bp of the 5' flanking region of the gene (Kassavetis *et al* ., 1989). Mutations introduced at conserved positions within the promoter, particularly in box B, dramatically reduce the affinity of TFIIIC for DNA (Fabrizio *et al* ., 1987; Newman *et al* ., 1983). Changes in the ionic environment can also affect transcription complex stability (Ruet *et al* ., 1984; Stillman *et al* ., 1984; Gabrielsen and Oyen, 1987).

The nucleotide sequence between the ICR of tRNA genes is generally considered to act as a spacer region, maintaining the optimum distance between the two parts of the promoter (Hofstetter *et al* ., 1981; Ciliberto *et al* ., 1982). The effects of varying the spacing between box A and box B upon the binding of transcription factors to the gene have been measured using mutant tRNA genes in which the spacing varied from 12-1530 bp (Ciliberto *et al* ., 1982; Dingerman *et al* ., 1983; Baker and Hall, 1984). A wide range of box A-B separations has been found to be compatible with tRNA gene transcription (Hall *et al* ., 1982; Carrara *et al* ., 1981; Raymond and Johnson, 1983; Dingermann *et al* ., 1983) but a distance of 34-53 nucleotides between boxes A and B appears to be the optimum for stable binding of TFIIIC. This range of distances allows for the efficient transcription of tRNA genes containing introns or sequences coding for a long variable arm, either of which would increase the spacing between box A and box B.

Recent research on the interactions of eukaryotic tRNA genes with their transcription factors has suggested the following model for the formation of the transcription complex (Geiduschek and Tocchini-Valentini, 1988; Dean and Berk, 1988; Schneider *et al* ., 1990; Kassavetis *et al* ., 1989 and 1990; Yoshinaga *et al* ., 1987 and 1989; Camier *et al* .,1990; Bartholomew *et al* ., 1990). Figure 1.2. shows a simplified summary of transcription factor binding onto tRNA genes. Firstly, TFIIIC/τ binds to the ICR, contacting box B first. The multiple subunits of this transcription factor allow TFIIIC/τ the flexibility needed for association with both parts of a split promoter. The binding of TFIIIC/τ to the gene allows TFIIIB to

become associated with the 5' end of the gene, and it is thought that this factor then sequesters RNA polymerase III onto the transcription complex. RNA polymerase III then rapidly cycles on and off the complex for multiple rounds of transcription. Transcription usually initiates at a specific purine nucleotide located between -20 and -3 bp upstream of the tRNA coding sequence (Clarkson, 1983) although some genes may have alternative initiation sites for transcription both *in vivo* and *in vitro* (Dingermann and Nerke, 1987; Thomann *et al* ., 1989; Schmutzler and Gross, 1990).

Recently it has been shown in yeast that once the transcription complex has been formed, TFIIIC/τ can be removed (by high salt or heparin treatment) without affecting further rounds of transcription (Kassavetis *et al* ., 1990). TFIIIB alone appears to be responsible for positioning RNA polymerase III for repeated cycles of transcription. TFIIIB is thus the sole transcription initiation factor in yeast, and TFIIIC/τ is purely an assembly factor for the transcription complex.

1.5.4. Termination of transcription.

The 3' ends of eukaryotic tRNA genes have a simple consensus sequence of four or more T residues that act as the signal to terminate transcription (Garber and Gage, 1979; Koski *et al* ., 1982; Koski and Clarkson, 1982). The run of Ts usually occurs within a few bp of the 3' end of the gene, although exceptions to this rule have been reported (Garber and Gage, 1979; Clarkson, 1983). Deletion of these residues causes read-through transcription (Adeniyi-Jones *et al* ., 1984; Allison and Hall, 1985), and mutations creating stretches of 4, 5, or 6 intragenic T residues cause premature termination of transcription (Koski *et al* ., 1980). In some cases shortening the run of T residues can affect termination, as in the case of a yeast $tRNA^{Tyr}$ gene where shortening T_7 to T_4 abolishes termination (Allison and Hall, 1985). A human $tRNA^{Met}$ gene produces multiple transcripts *in vitro* as a result of a series of inefficient termination sites of less than four T residues (Vnencak-Jones *et al* ., 1987). Other short sequences have been shown to act as termination signals. The sequences ATCTT and CTTCTT were demonstrated to act as termination signals for human $tRNA^{Val}$ genes (Thomann *et al* ., 1989).

1.5.5. Processing of tRNA gene transcripts.

Transcription of tRNA genes yields a precursor tRNA molecule that has to be processed to form the mature tRNA (for review see Deutscher, 1984). Processing occurs in a stepwise fashion, and involves endonucleolytic removal of the 5' leader and the 3' trailer, addition of the 3' terminal CCA, site-specific base modifications and if present, excision of the intron (Sharp *et al* ., 1985; Dingermann and Nerke, 1987; Thomann *et al* ., 1989; Marchfelder *et al* ., 1990; O'Connor and Peebles, 1991). The order of the steps involved in maturation does not appear to be the same for all tRNAs; different precursor tRNAs have been shown to be processed along different pathways (Rooney and Harding, 1986; van Tol *et al* ., 1987) and alternative processing paths for one tRNA may occur *in vivo* (O'Connor and Peebles, 1991). The enzymes involved in processing recognise the universally conserved structural features of tRNAs, which allows all tRNA precursors to be processed by the same apparatus (Castagnoli *et al* ., 1982; Mattocia *et al* ., 1983; Swerdlow and Guthrie, 1984). It has been shown that chemical modification of conserved nucleotides can inhibit the processing of pre-tRNAs (Thurlow *et al* ., 1991). After the completion of processing the mature tRNA leaves the nucleus.

It is uncertain why tRNAs are transcribed as precursor molecules. Possibly the presence of extra nucleotides in such precursors prevents immature transcripts from leaving the nucleus (De Robertis *et al* ., 1981; Melton and Cortese, 1979). Maturation of the 3' end of the precursor is presumably necessary to remove the termination signal. The exact role of the specific base modifications characteristic of mature tRNAs is also poorly understood, although some individual examples have been studied. It is known that all mature eukaryotic tyrosine tRNAs have a pseudouridine (Ψ_{35}) in the central position of the anticodon and that modification of this residue is important in codon-anticodon interactions (van Tol and Beier, 1988; Choffat *et al* ., 1988).

It has been shown for a yeast $tRNA^{Tyr}$ processed in *Xenopus* oocytes that the 5' leader is removed in three stages, the last of which is apparently accompanied by excision of the 3' trailer and addition of the 3' CCA (De Robertis and Olsen, 1979; Melton *et al* ., 1980; De Robertis *et al* ., 1981). In *in vitro* transcription reactions this three stage process has not generally been observed. The CCA motif occurs at the 3' end of all mature tRNAs, but is not coded for in any eukaryotic tRNA gene.

The three bases are added post-transcriptionally by the enzyme tRNA nucleotidyl transferase.

The $tRNA^{Tyr}$ gene family is one of at least ten tRNA gene families in yeast which contain intervening sequences (Ogden *et al* ., 1984; Stucka and Feldman, 1988). The presence of an intron in tRNA genes does not appear to be related to transcriptional activity. However, the intron present in all yeast $tRNA^{Tyr}$ genes has been shown to be essential for modification of its tRNA product (Johnson and Abelson, 1983; Choffat *et al* ., 1988). The same requirement for an intron occurs for human $tRNA^{Tyr}$ genes (van Tol and Beier,1988). In higher eukaryotes, only a small proportion of the genes described so far contain an intervening sequence. Apart from the human $tRNA^{Tyr}$ genes mentioned above, introns have only been found in members of the human $tRNA^{Leu}$ (Green *et al* ., 1990) and $tRNA^{Arg}$ (see Chapter three and Chapter four) gene families. Where introns are present, neither their sequence nor their length appear to be conserved between different tRNA genes but the position of the intron is always the same, one nucleotide 3' of the anticodon. Within gene families, the sequence and length of introns tends to be conserved (Szekely *et al* ., 1988). The intron is removed by a two stage process, excision of the intron followed by splicing of the two tRNA half molecules (Abelson, 1979; O'Farrell *et al* ., 1978; Knapp *et al* ., 1978; Standring *et al* ., 1981; De Robertis *et al* ., 1981; Willis *et al* ., 1984; Baldi *et al* ., 1986). As for the other steps in tRNA processing, the mechanism relies on the conformation of the molecule rather than any sequence motif: point mutations which affect the secondary stucture of an *Arabidopsis* $tRNA^{Tyr}$ gene cause inefficient splicing of intron-containing pre-tRNAs (Szweykowska-Kulinska and Beier, 1991).

1.5.6. The role of flanking sequences in transcription modulation.

As described in sections 1.5.1. and 1.5.2. the transcription of tRNA genes is predominantly controlled by an intragenic split promoter. However it has now been well established that flanking sequences modulate transcription in *B. mori* (Sprague *et al* ., 1980; Larson *et al* ., 1983; Morton and Sprague, 1984; Young *et al* ., 1986), in *S. cerevisiae* (Raymond and Johnson, 1983 and 1987; Johnson and Raymond, 1984; Shaw and Olsen, 1984), in *D. melanogaster* (Dingermann *et al* ., 1982;

Schaack *et al* ., 1984; Sajjadi *et al* ., 1987), in *Xenopus* (Hipskind and Clarkson, 1983) and in humans (Arnold and Gross, 1987; Morry and Harding, 1986). Genes with identical coding sequences but different flanking sequences have been found to differ in their transcriptional efficiency (Lofquist and Sharp, 1986; Gouilloud and Clarkson, 1986; Arnold *et al* .,1986 and 1988; Doran *et al* ., 1987 and 1988). Flanking regions are not generally conserved between different tRNA gene copies, and so far no definite mechanism for modulation by extragenic sequences has been established. It is thought that tissue- or developmental stage-specific factors may interact with control elements in the flanking sequences to regulate gene expression. There may be several levels of control over gene expression; the methylation status of bases flanking tRNA genes can have an effect on transcription (Besser *et al* ., 1990) and recently it has been shown that changes in the levels of selinium in culture can affect the levels of expression of selenocysteine $tRNA^{[Ser]Sec}$ genes in mammalian cells (Hatfield *et al* ., 1991).

The majority of experiments on the role of flanking sequences have isolated elements in the 5' flank of tRNA genes, although changes in the 3' flank have also been shown to affect transcription (Wilson *et al* ., 1985; Wu *et al* ., 1987; Schaack *et al* ., 1983; Arnold *et al* ., 1988). Most such studies have been carried out *in vitro* with some exceptions (Shaw and Olsen, 1984; Schaack and Soll, 1985; Raymond *et al* ., 1985; Marschalek and Dingermann, 1988). The predominant finding has been that deletions of the 5' flank diminish transcription to a greater or lesser extent. However, in some cases the replacement of the 5' flank with unrelated DNA actually increases transcription, suggesting that negative as well as positive modulator elements exist (Arnold *et al* ., 1988). These results also suggest that this effect may not be very sequence specific.

Positive modulators have been isolated from the 5' flanks of a number of tRNA genes (Sprague *et al* ., 1980; Larson *et al* ., 1983; Shaw and Olsen, 1984; Johnson and Raymond, 1984; Raymond *et al* ., 1985; Morry and Harding, 1986; Sajjadi and Spiegelman, 1987; Arnold and Gross, 1987; Arnold *et al* ., 1988; Lofquist *et al* ., 1988; Horvath and Spiegelman, 1988; Rooney and Harding, 1988). These elements are of 5-35 nucleotides in length, and occur within about 50 bases upstream of the gene (although a tRNA-like structure extending further upstream of a human $tRNA^{Glu}$ gene appears to affect its transcription; Goddard *et al* ., 1983; Gonos and Goddard, 1990b). Potential Z-DNA sequences in the 5' flanks of tRNA genes appear to inhibit their transcription (Hipskind and Clarkson, 1983; Nordheim

and Rich, 1983; Santoro *et al* ., 1984; Glew *et al* ., 1986), as does an undecanucleotide sequence in the 5' flanking region of a *Drosophila* $tRNA^{Lys}$ gene (De Franco *et al* ., 1981).Generally when modulatory elements exist, their position relative to the coding sequence is of importance in determining their effect.

Finally, it has been demonstrated in several cases that 5' flanking sequences modulate transcription differently in different transcription assays, and when using extracts from different sources (Raymond and Johnson, 1983; Johnson *et al* ., 1984; Schaack and Soll, 1985; Drabkin and RajBhandry, 1985a and 1985b; Gonos and Goddard, 1990c; Schmutzler and Gross, 1990) which supports the idea that their role *in vivo* may be in controlling tissue-specific expression.

CHAPTER TWO

MATERIALS AND METHODS.

2.1. Materials.

2.1.1. List of suppliers.

Unless otherwise specified all chemicals were Analar grade, supplied by BDH Chemicals, Poole, Dorset, or Fisons Scientific, Loughborough, Leics. All radiochemicals were provided by Amersham International plc. Growth media were supplied by Difco Laboratories, Detroit, USA. Where chemicals or equipment were obtained from other sources, this is indicated in the text. A list of suppliers of special reagents is given below:

Aldrich Chemicals Co., Gillingham, Dorset
Amersham International plc., Amersham, Bucks.
Anglian Biotechnology Ltd., Colchester
Boehringer Corporation (London) Ltd., Lewes, E. Sussex
Fuji Photofilm Co., Japan
Gibco/BRL (Bethesda Research Laboratories) Ltd., Paisley
IBI Ltd., Cambridge
Kodak Ltd., Kirby, Liverpool
Northumbria Biologicals Ltd., Cramlington, Northumberland
Pharmacia Ltd., Milton Keynes
Promega Ltd., Southampton
Sigma Chemicals Co., Poole, Dorset

2.2. Media and buffers.

All media and solutions used for the growth of microorganisms or the handling

of nucleic acids were sterilised as appropriate by autoclaving for 20 minutes at 15 psi or by filter sterilisation.

2.2.1. Bacterial growth media.

LB medium (per litre) : 10 g Bacto tryptone
5 g Yeast extract
10 g NaCl

LB plates (per litre) : 1 litre LB
15 g Bacto agar

LBA plates (per litre) : 996 ml LB
15 g Bacto agar
4 ml Ampicillin (25 mg/ml stock solution)

LB top agar (per litre) : 1 litre LB
7g Bacto agar

2 x TY (per litre) : 16 g Bacto tryptone
10 g Yeast extract
5 g NaCl

Minimal glucose plates (per litre) : 15 g minimal agar in 900 ml H_2O
100 ml 10 x M9 salt
1 ml 0.1 M $CaCl_2$
10 ml 20% (w/v) glucose
1 ml 1M $MgSO_4$
1 ml 1M thiamine HCl

H-plates (per litre) : 10 g Bacto tryptone
8 g NaCl
8 g agar

H-top agar (per litre) : 10 g Bacto tryptone
8 g NaCl
8 g agar

2.2.2. Buffers.

PBS mix : 8/1/1 mix of solutions A, B and C
Solution A : 170 mM NaCl
0.5 mM KCl
10 mM $Na_2 HPO_4$
2 mM $KH_2 PO_4$
Solution B : 7 mM $CaCl_2.2H_2O$
Solution C : 5 mM $MgCl_2.6H_2O$

10 x TBE : 0.89 M Tris
0.89 M boric acid
0.025 M EDTA
(this gives pH 8.3)

20 x SSC : 3 M NaCl
0.3 M tri-sodium citrate

20 x SSPE : 3.6 M NaCl
0.2 M sodium phosphate, pH 7.7
0.02 M EDTA

TE : 10 mM Tris-HCl
1 mM EDTA
Adjusted to pH 7.6 with HCl

STET buffer : 50 mM Tris-HCl pH 8.0
50 mM EDTA
8% (w/v) sucrose
5% (v/v)Triton x-100

Quiagen plasmid DNA preparation

Buffer A : 400 mM NaCl
50 mM MOPS
15% ethanol
Adjusted to pH 7.0

Buffer C : 1 M NaCl
50 mM MOPS
15% ethanol
Adjusted to pH 7.0

Buffer F : 1.5 M NaCl
50 mM MOPS
15% ethanol
Adjusted to pH 7.5

Human genomic DNA preparation

Blood lysis buffer : 0.32 M Sucrose
10 mM Tris-HCl pH 7.5
5 mM $MgCl_2$
1% Triton X-100

Nuclear lysis buffer : 10 mM Tris-HCl pH 8.2
0.4 M NaCl
2 mM EDTA

2.3. Bacterial strains and storage.

Bacterial strains were stored in the short term (for up to one month) at 4 °C on tightly sealed LB, H- or minimal glucose plates. Longer term storage was in 50% (v/v) glycerol at -20 °C. The following *E. coli* strains were used in this study.

A. Strain used as host to plasmid pUC19 and bacteriophage M13 recombinants:
TG1 *supE hsd* Δ5 *thi* Δ(*lac-proAB*) F' [*tra D* 36 *proAB* + *lacI* q *lacZ* ΔM15] (Gibson, 1984).

B. Strain used for bacteriophage λ recombinants:
BHB 2600 r^-_k m^+_k *SupE* + *SupF* +*met* - (Hohn, 1979).

2.4. DNA vectors, recombinants and oligonucleotides used in this study.

DNA vectors, recombinants and oligonucleotides were stored in TE (see section 2.2.2.) at -20 °C. DNA stored in this way remains stable indefinitely.

2.4.1. λHt363.

The recombinant bacteriophage λHt363 is one of several λ Charon 4A recombinants selected from a human genomic DNA library by Dr. J.P. Goddard using a mixed tRNA probe labelled by ligation of [^{32}P]pCp (see section 2.5.11.4). This recombinant had been partially characterised as described in section 3.1.

2.4.2. M13 bacteriophage vectors.

The bacteriophage vectors M13mp10, M13mp11, M13mp18 and M13mp19 (Norrander *et al* ., 1983; Yanisch-Perron *et al* ., 1985) were used to subclone DNA fragments for sequencing by the chain termination method (Sanger *et al* ., 1980). M13mp18 and M13mp19 vectors are shown in figure 2.1. M13mp10 and M13mp11 differ from M13mp18 and M13mp19 only in the smaller number of restriction sites in the polylinker multiple cloning site.

2.4.3. pUC19 plasmid vector.

The plasmid vector pUC19 (Yanisch-Perron *et al* .,1985) was used for all of the subcloning of restriction fragments from λHt363 described in this study, other than subclones used for sequencing purposes. All of the pUC19 recombinants described were constructed by inserting restriction fragments into the multiple cloning site. The pUC19 vector is shown in figure 2.2.

2.4.4. Synthetic oligonucleotides.

Synthetic oligonucleotides (for use as primers in the polymerase chain reaction; see section 2.5.15. and table 4.1.) were made in the Department of Biochemistry by Dr. V. Math, using an Applied Biosystems 381A DNA synthesiser for the phosphite-triester method (Atkinson and Smith, 1984).

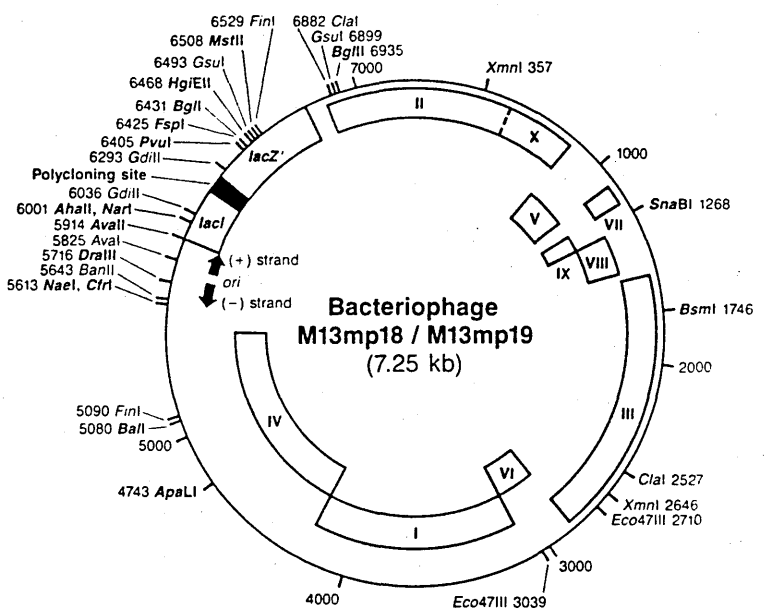
2.5. Methods.

2.5.1. Small scale preparation of plasmid DNA.

The method used was an adaptation of that of Holmes and Quigley (1981). A 3 ml overnight culture was prepared from a single colony of transformed bacteria in LB (see section 2.3.) supplemented with the appropriate antibiotic (100 μg/ml ampicillin for pUC19 and recombinants derived from pUC19). 1.5 ml of the

Figure 2.1. Restriction map of bacteriophage M13 vectors M13mp18 and M13mp19.

The bacteriophage M13 vectors M13mp10, M13mp11, M13mp18 and M13mp19 were used in this study for nucleotide sequence analysis by the Sanger chain termination method. The polycloning sites for M13mp18 and M13mp19 are shown below. M13mp10 and M13mp11 differ from M13mp18 and M13mp19 only in the lack of the *Kpn* I and *Sph* I restriction enzyme sites in the polycloning site.
(From Sambrook *et al* ., 1989).



Polycloning Sites
M13mp18

	1	2	3	4	5	6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	7	8
	Thr	Met	Ile	Thr	Asn	Ser	Ser	Ser	Val	Pro	Gly	Asp	Pro	Leu	Glu	Ser	Thr	Cys	Arg	His	Ala	Ser	Leu	Ala	Leu	Ala
ATG	ACC	ATG	ATT	ACG	AAT	TCG	AGC	TCG	GTA	CCC	GGG	GAT	CCT	CTA	GAG	TCG	ACC	TGC	AGG	CAT	GCA	AGC	TTG	GCA	CTG	GCC

EcoRI *SacI* *KpnI* *SmaI* *XmaI* *BamHI* *XbaI* *SalI* *AccI* *HincII* *PstI* *SphI* *HindIII*

M13mp19

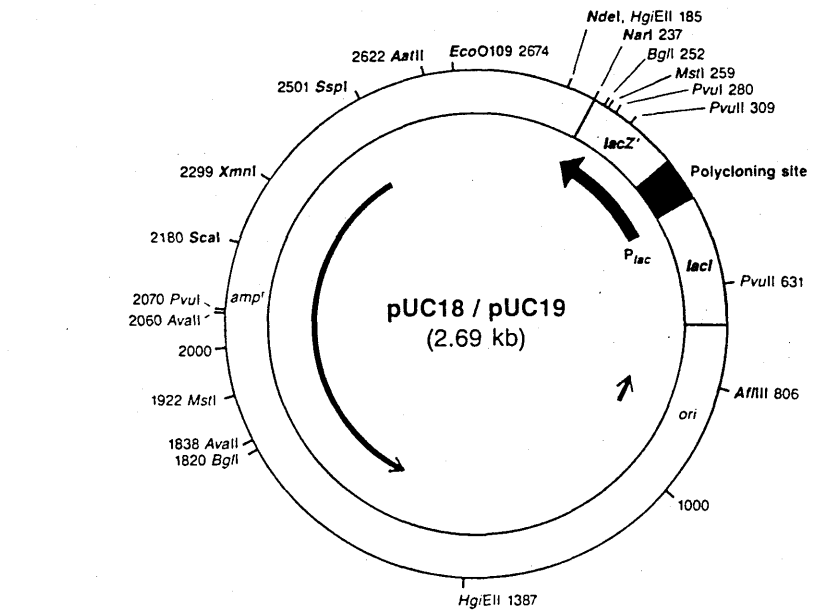
	1	2	3	4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	5	6	7	8
	Thr	Met	Ile	Thr	Pro	Ser	Leu	His	Ala	Cys	Arg	Ser	Thr	Leu	Glu	Asp	Pro	Arg	Val	Pro	Ser	Ser	Asn	Ser	Leu	Ala
ATG	ACC	ATG	ATT	ACG	CCA	AGC	TTG	CAT	GCC	TGC	AGG	TCG	ACT	CTA	GAG	GAT	CCC	CGG	GTA	CCG	AGC	TCG	AAT	TCA	CTG	GCC

HindIII *SphI* *PstI* *SalI* *AccI* *HincII* *XbaI* *BamHI* *SmaI* *XmaI* *KpnI* *SacI* *EcoRI*

In M13mp18, the *EcoRI* site lies immediately downstream from P_{lac}
In M13mp19, the *HindIII* site lies immediately downstream from P_{lac}

Figure 2.2. Restriction map of the plasmid vectors pUC18 and pUC19.

The plasmid vector pUC19 was used for all of the subcloning of DNA fragments described in this study, apart from subclones needed for nucleotide sequence analysis. This figure also shows the polycloning site of pUC18, which differs from that of pUC19 only in its orientation.
(From Sambrook *et al* ., 1989).



Polycloning Sites

pUC18

	1	2	3	4	5	6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	7	8
	Thr	Met	Ile	Thr	Asn	Ser	Ser	Ser	Val	Pro	Gly	Asp	Pro	Leu	Glu	Ser	Thr	Cys	Arg	His	Ala	Ser	Leu	Ala	Leu	Ala
ATG	ACC	ATG	ATT	ACG	AAT	TCG	AGC	TCG	GTA	CCC	GGG	GAT	CCT	CTA	GAG	TCG	ACC	TGC	AGG	CAT	GCA	AGC	TTG	GCA	CTG	GCC

EcoRI SacI KpnI SmaI XmaI BamHI XbaI SalI AccI HincII PstI SphI HindIII

pUC19

	1	2	3	4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	5	6	7	8
	Thr	Met	Ile	Thr	Pro	Ser	Leu	His	Ala	Cys	Arg	Ser	Thr	Leu	Glu	Asp	Pro	Arg	Val	Pro	Ser	Ser	Asn	Ser	Leu	Ala
ATG	ACC	ATG	ATT	ACG	CCA	AGC	TTG	CAT	GCC	TGC	AGG	TCG	ACT	CTA	GAG	GAT	CCC	CGG	GTA	CCG	AGC	TCG	AAT	TCA	CTG	GCC

HindIII SphI PstI SalI AccI HincII XbaI BamHI SmaI XmaI KpnI SacI EcoRI

In pUC18, the *Eco*RI site lies immediately downstream from P_{lac}
In pUC19, the *Hind*III site lies immediately downstream from P_{lac}

overnight culture was transferred to an Eppendorf tube and the bacteria harvested by centifugation at 10,000 rpm for 5 min. The remainder of the overnight culture was set aside for the preparation of a glycerol stock if necessary (see section 2.4.1.). The bacterial pellet was dried and resuspended in 700 µl of cold STET buffer (see sectin 2.2.2.). 50 µl of a solution of fresh cold lysozyme (10 mg/ml) was added and the solution was left on ice for 5 min. The samples were then heated to 95 °C for one minute and bacterial debris removed by centrifugation at 10,000 rpm for 15 min. The bacterial lysate was extracted twice with an equal volume of a 1:1 mixture of TE saturated phenol:chloroform (see section 2.5.5.1.). After the addition of 50 µl 3M sodium acetate and 400 µl isopropanol, and a period of 20 min at -20 °C plasmid DNA was precipitated by centrifugation (10,000 rpm for 15 min at 4 °C). The plasmid DNA was washed once with 70% ethanol and once with 100% ethanol, dried at room temperature and redissolved in 50 µl TE. A 10 µl sample from each preparation was subjected to restriction endonuclease digestion and electrophoresed through an agarose gel to check the size and quality of the plasmid DNA.

2.5.2. Plasmid DNA preparation using QUIAGEN pack cartridges.

This method was used to prepare up to 100 µg of high purity plasmid DNA. The method is that described in "The Quiagenologist" Application Protocols (Diagen GmbH). Transformed cells from a 50 ml overnight culture prepared from a single colony (in LB+ 100 µg/ml ampicillin) were harvested by centrifugation at 2,500 rpm for 5 min at 4 °C. The cells were resuspended in 3.6 ml of an ice-cold solution of 50 mM Tris-HCl, pH 7.4. To this suspension 0.4 ml of a fresh solution of 20 mg/ml lysozyme was added and the cells left at 0 °C. After 10 min 1 ml of 0.5 M EDTA (pH 8.0) was added, and after a further 10 min at 0 °C 0.2 ml of 2% Triton X-100. The mixture was left on ice for 1 h then the cell debris and chromosomal DNA removed by centrifugation at 10,000 rpm for 45 min at 4 °C. The cleared lysate was treated with RNase A (final concentration 20 µg/ml) for 30 min at 37 °C then proteinase K (final concentration 10 µg/ml) again for 30 min at 37 °C. The treated lysate was centrifuged at 10,000 rpm for 10 min at 4 °C. 5 ml of the supernatant was added to 1.1 ml 5M NaCl and 0.5 ml 1M MOPS, pH 7.0. A

Quiagen pack cartridge was equilibrated with 2 ml of buffer A (see section 2.2.2.) then the DNA adsorbed onto the cartridge by forcing the sample through using a sterile 5 ml syringe. Proteins, RNA, phenol and any other contaminants were removed by washing the cartridge with 5 ml of buffer C (see section 2.2.2.). The plasmid DNA was eluted from the cartridge using 2 ml of buffer F (see section 2.2.2.), and precipitated by the addition of 0.8 vol. of isopropanol. After 30 min on ice samples were centrifuged at 10,000 rpm for 15 min at 4 °C. The DNA pellet was washed once with 70% ethanol, once with 100% ethanol and then dried at room temperature.The concentration of the DNA was determined by measuring the A_{260}, using the assumption that a 50 µg/µl solution of DNA has an A_{260} of 1.0 when measured in a spectrophtometer cell with a 1 cm light-path. The A_{280} of the DNA preparation was also determined, since the ratio A_{260}:A_{280} gives an indication of the purity of the DNA. Pure double-stranded DNA gives an A_{260}:A_{280} of 1.8:1. Finally, a sample of each preparation was subjected to restriction endonuclease digestion and together with an uncut sample electrophoresed on an agarose gel to check the quality of the preparation.

2.5.3. Large scale preparation of bacteriophage λ DNA.

A single colony of the bacterial host *E. coli* BHB 2600 was inoculated into a 10 ml overnight culture of LB. A suitable dilution of phage, giving 10-100 plaques, was adsorbed onto 200 µl of the overnight culture of the bacterial host, mixed with 3 ml LB top agar (see section 2.2.1.), poured onto a LB plate (see section 2.2.1.) and incubated overnight. A single plaque was removed from the plate, added to 200 µl of an overnight culture of bacterial host and left standing for 20 min at room temperature. The culture was then transferred to a fresh flask containing 20 ml LB supplemented with 10 mM $MgSO_4$ and 0.2% (w/v) maltose. The flask was shaken at 37 °C until lysis occured (usually between 4 to 6 h after inoculation), when 1 ml of chloroform was added. After 10 min of shaking, the upper layer was sedimented by centrifugation at 2,000 rpm for 20 min and the supernatant stored at 0 °C in a fresh tube. The titre of the supernatant was generally about 10^{10} pfu/ml. Two 500 ml batches of LB, supplemented with 10 mM $MgSO_4$ and 0.2% maltose, were inoculated with 2 x 5 ml of an overnight bacterial culture. The culture was grown

until A_{610} reached 0.3 when a total of 5 x 10^{10} pfu of phage was added. (This gives a ratio of pfu:host cells of 1:200). The infected culture was shaken until lysis occured (usually between 3 and 4 h) and then 5 ml of chloroform was added to each flask and shaking continued for a further 10-15 min.The lysed cultures were decanted and bacterial debris was removed by centrifugation at 4,000 rpm for 20 min. DNaseI and pancreatic RNase were added (each to a final concentration of 10 µg/ml) to the supernatant which was incubated at room temperature for 30 min. Solid NaCl was added to 2%, solid polyethylene glycol (PEG-6000 Serva) was added to 8% and the culture was left standing at 4 °C overnight to precipitate the phage particles. The phage was sedimented by centrifugation at 6,000 rpm for 30 min and carefully resuspended in 20 ml of lambda diluent (100 mM Tris-HCl, pH 8; 1mM EDTA ; 10 mM Mg SO_4) to which 0.71 g of solid CsCl was added per ml of sample. After a clarifying centrifugation at 1,500 rpm for 30 min the solution was transferred into sealable tubes and centrifuged at 48,000 rpm and 20 °C overnight using a VTi50 rotor of a Beckman ultracentrifuge. The white phage band was collected after piercing the tube with a 18 g needle just below the band.The phage was then dialysed against three changes (1 h each) of 500 ml buffer (10 mM Tris-HCl, pH 7.5; 1mM EDTA ; 10 mM Mg SO_4) to remove CsCl. DNA was extracted with 1:1 phenol/chloroform (see section 2.5.5.1.) twice, and ether twice. After the addition of 2.5 vol. ethanol the DNA was precipitated by centrifugation at 10,000 rpm for 10 min, dried and resuspended in 0.5 ml TE. The DNA concentration was determined as described in section 2.5.2.

2.5.4. Preparation of high molecular weight human genomic DNA from blood.

The method used is that described by Miller *et al* ., (1988). 40 ml of blood lysis buffer (see section 2.2.2.) was added to 10 ml of fresh or frozen whole blood kept on ice. The mixture was centrifuged at 2,800 rpm for 10 min at 4 °C. The cell pellet was then resuspended in 3 ml of nuclear lysis buffer (see section 2.2.2.). To the suspension 0.2 ml 10% SDS and 0.1 ml proteinase K (10 mg/ml in 1% SDS, 2 mM EDTA) were added and the mixture incubated at 37 °C overnight. After the addition of 1 ml of a saturated solution of NaCl (approximately 6 M) the mixture

was extracted gently with 1:1 phenol/chloroform two or three times until the interface was clear. Two volumes of ethanol were added to the aqueous phase and the precipitated DNA spooled out onto a closed Pasteur pipette. The DNA was washed once in 70% ethanol and once in 100% ethanol and redissolved in an appropriate volume of TE. The concentration of the DNA was determined as described in section 2.5.2. Agarose gel electrophoresis of a sample of the DNA confirmed that genomic DNA prepared in this way was over 50 kb in size. Genomic DNA was stored at 4 °C.

2.5.5. Extraction and precipitation of DNA.

2.5.5.1. Phenol/chloroform extraction.

Phenol/chloroform stock was made by mixing equal amounts of phenol and chloroform then equilibrating the mixture by extracting several times with TE. The equilibrated mixture was stored under TE in the dark. 8-Hydroxyquinoline was added to 0.1% as an antioxidant and to assist in distinguishing the phenol/chloroform from the aqueous phase in extractions. Extraction with phenol/chloroform was performed as follows: an approximately equal volume of phenol was added to the DNA solution to be extracted, mixed and centrifuged for 5 min; the upper aqueous layer was transferred to a fresh tube and the extraction repeated as necessary.

2.5.5.2. Ethanol and isopropanol precipitation.

Ethanol used was of the absolute alcohol 100 grade and was stored at -20 °C. A salt solution was added to the DNA solution to the appropriate concentration (0.1 volume of 5 M ammonium acetate or 0.1 volume of 3 M sodium acetate) and mixed with 2.5 volumes of cold ethanol. The samples were placed at -20 °C or in dry ice for an appropriate period then centrifuged at 10,000 rpm for 15 min at 4 °C. The

precipitated DNA was washed successively with 70% ethanol and 100% ethanol and dried at room temperature under vacuum. DNA was usually stored at -20 °C in TE. When the DNA sample volume was such that addition of 2.5 volumes of ethanol was impractical, precipitation was carried out by the addition of an equal volume of isopropanol to the DNA + salt solution.

2.5.6. DNA digestion with restriction endonucleases.

Restriction digests were routinely carried out in 1.5 ml Eppendorf tubes in the presence of the appropriate buffer. One unit of restriction enzyme activity is defined as the amount of enzyme required to digest 1 μg of λ DNA to completion in one hour at the correct incubation temperature and in the appropriate buffer. In practice a several fold excess of enzyme was usually added to digest the DNA. A typical reaction mixture contained 1 μg of DNA, the correct restriction enzyme buffer (usually supplied by the manufacturer) and 5 units of the desired restriction enzyme in a final volume of 25 μl. The mixture was incubated for at least one h at the temperature recommended by the manufacturer, usually 37 °C. The completion of the digestion was monitored by electrophoresis of an aliquot through an agarose gel. After digests, restriction enzymes were heat inactivated (65 °C for 10 min) or removed by phenol/chloroform extraction. Digestions of genomic DNA were incubated overnight, using 8 units of enzyme per μg of genomic DNA. For multiple digestions where different buffers were required, the digestion which required the lower ionic strength was carried out first, then the ionic conditions were adjusted, the second enzyme added and the incubation continued.

2.5.7. Separation of DNA fragments by electrophoresis.

2.5.7.1. DNA electrophoresis in agarose and polyacrylamide gels.

Agarose gels were routinely used to separate DNA fragments in the size range

0.1-20 kb. They were prepared by dissolving agarose (BRL ultra PURE) in 1 x TBE (see section 2.2.2.) to the desired concentration (0.5-2%; Sambrook *et al* ., 1989). Ethidium bromide was added to 10 µg/ml and the gel poured onto a horizontal tank (BRL model H5 or model 200). Wells with a suitable capacity were formed by insertion of a comb into the gelling agarose. After submerging the gel in 1 x TBE (see section 2.2.2.) one fifth volume of loading buffer (10% glycerol, 25% ficoll 400, 1 mM EDTA and 0.25% bromophenol blue) was added to the DNA samples and the mixture loaded into the wells. Electrophoresis time and current (20-50 mA) were varied as needed. After electrophoresis gels were destained by two x 10 min washes in 1 x TBE or distilled water. DNA fragments were visualised by the use of a long wavelength u.v. transilluminator (U.V. Products Inc.) and photographed with a Polaroid Cu-5 camera, using type 665 positive/negative film. The sizes of the restriction fragments were determined by comparison with DNA marker fragments of known size, electrophoresed on the same gel. The distances (in mm) travelled by the known marker fragments were plotted on semi-log graph paper against log size of DNA (kb). The sizes of unknown fragments could then be determined by measuring the distance travelled and referring to the standard curve. When ^{32}P or ^{35}S labelled DNA fragments were separated on agarose gels the gels were dried under vacuum on a gel drier (model 1125B Bio-Rad, 2 h at 80 ^{o}C) prior to autoradiography.

Polyacrylamide gels were used to separate DNA fragments of 0.2 kb and below. Gels were prepared as follows: a deionised 40% stock solution (38% acrylamide, 2% N'N'-methylene bisacrylamide) was diluted to the desired concentration (usually 10%) in 1 x TBE. After the addition of 300 µl of fresh 10% ammonium persulphate and 50 µl TEMED, the mixture was poured between two glass plates and a comb placed at the top of the gel. After polymerisation, DNA samples plus loading buffer were placed in the wells and electrophoresis carried out at 40 mA using a vertical gel kit (BRL model V161).

2.5.7.2. Recovery of DNA from low melting point agarose.

Low melting point agarose (BRL ultra PURE) gels were prepared and used in the same way as standard agarose gels. The DNA band of interest was visualised and

the gel containing it excised over u.v. and transferred to an Eppendorf tube. The gel sample was placed at 70 °C in a waterbath for 8 min to melt the agarose and three volumes of TE were added. The sample was then extracted twice with phenol saturated with TE and three times with ether to remove the agarose and the DNA was then precipitated with ethanol.

2.5.8. Subcloning DNA fragments into pUC19.

Insert DNA and pUC19 vector DNA were digested with the appropriate enzyme(s) and the extent of digestion checked by gel electrophoresis. Insert DNAs were usually purified from low melting point agarose gels as described in section 2.5.7.2. In some cases, to reduce background due to religation of the vector the digested pUC19 DNA was treated with alkaline phosphatase to remove 5' phosphates: the reaction mix (50 µl), containing 5 µg digested vector DNA, 1 unit of calf intestine alkaline phosphatase (Boehringer grade I), 50 mM Tris-HCl (pH 9), 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$ and 1 mM spermidine was incubated at 37 °C for 1 h. After heat inactivation of the enzyme (65 °C for 10 min) the digested, dephosphorylated vector DNA was purified by phenol/chloroform extraction followed by ethanol precipitation and resuspended at 20 ng/µl in TE. Typical ligation reactions contained: insert DNA (20-100 ng), vector DNA (20-50 ng), 1 mM ATP, 10 mM DTT, 50 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, 2 mM spermidine, bovine serum albumin (100 µg/ml) and 2 units of T4 DNA ligase in a final volume of 10 µl. The ligation mixture was incubated at 15 °C for 16 h. Greater concentrations of T4 ligase (5-10 units in 10 µl) were used for the ligation of DNA molecules with blunt ends.

2.5.9. Transformation of bacterial cells with plasmid DNA.

Competent *E. coli* cells were prepared by a modified version of the method of Dagert and Erlich (1979). 2 ml of an overnight 10 ml LB culture of *E. coli* TG1 was used to inoculate 40 ml of LB. The cells were grown at 37 °C with shaking for

2-3 h until the A_{550} of the culture exceeded 0.3. The cells were cooled on ice and harvested by centrifugation at 2,500 rpm for 10 min at 4 °C. The cells were gently resuspended in 20 ml of ice cold sterile 50 mM $CaCl_2$ and left on ice for 20 min. The cells were pelleted as before and resuspended in 4 ml sterile 50 mM $CaCl_2$. Competent cells prepared in this way remained usable for up to 5 days if stored on ice. Transformations with pUC19 DNA were carried out in Eppendorf tubes. 5 µl of ligation mix containing 25-50 ng of DNA was added to 100 µl of competent cells and the mixture incubated on ice for at least 40 min, then heat shocked at 42 °C for 1 min. The cells were placed on ice then spread together with 50 µl of 100 mM IPTG and 50 µl of X-gal (2% (w/v) in dimethylformamide) onto LB plates containing 100 µg/ml ampicillin. Plates were incubated at 37 °C overnight to allow colony formation. Colonies transformed with recombinant pUC19 were colourless in the presence of X-gal and IPTG because they lacked a functional β-galactosidase. Colonies transformed with pUC19 alone are blue in the presence of X-gal and IPTG.

2.5.10. DNA transfer to membranes (Southern blotting).

2.5.10.1. DNA transfer to nylon membranes.

This method is basically that of Southern (1975). After electrophoresis (see section 2.5.7.1.) the agarose gel was immersed in 0.25 M HCl for 15 min at room temperature with gentle agitation. (This "depurination" step was only necessary to increase the efficiency of transfer of DNA fragments of over 10 kb). The gel was rinsed in distilled water then immersed in denaturation buffer (1.5 M NaCl, 0.5 M NaOH) for 30 min at room temperature. After rinsing the gel was washed in neutralising solution (1.5 M NaCl, 0.5 M Tris-HCl pH 7.2, 1 mM EDTA) for 2 x 15 min at room temperature. Blots were set up as follows: 10 x SSC (see section 2.2.2.) was used to fill two buffer reservoirs connected over a solid support by two wicks of Whatman 3 MM paper; the gel was placed on the bridge. A cling film sheet covered the whole apparatus, leaving a window of the same size and position of the gel, to prevent side capillary transfer. A nylon filter (Hybond-N, Amersham

International plc) cut to the size of the gel was placed on the gel and four sheets of 3 MM paper cut to the same size placed on top. An excess stack of absorbent paper was further placed on top and the whole system compressed using a 1 kg weight on top of a glass plate. Transfer of DNA was allowed to proceed for 16-20 h at room temperature. After blotting the nylon filter was marked for orientation, rinsed in 2 x SSC then baked at 80 °C for 5 h in a vacuum oven to bind the DNA to the membrane. Membranes were stored in sealed plastic bags at room temperature.

2.5.10.2. DNA transfer to nitrocellulose membranes.

The method used was that described for nylon membranes (section 2.5.10.1.) with the following variations. The denaturation buffer used was 0.2 M NaOH, 0.6 M NaCl and the neutralisation buffer was 1 M Tris-HCl (pH 7.5), 0.6 M NaCl. Nitrocellulose membranes (Schleicher and Schuell) were blotted and stored in the same way as described in section 2.5.10.1.

2.5.11. Radiolabelling of nucleic acids.

2.5.11.1. Random oligonucleotide-primed synthesis from DNA fragments.

The method used was that described by Feinberg and Vogelstein (1983 and 1984). DNA fragments were separated on a low melting point agarose gel as described in section 2.5.7.2. Two volumes of distilled water were added to the gel slice containing the excised band and the sample was heated to 95 °C for five minutes before being transferred to a 37 °C waterbath. A volume of diluted gel containing approximately 100 ng of DNA was added to the following: 10 µl of oligolabelling buffer (see below), 20 µg of bovine serum albumin, 30 µCi of α-[^{32}P]-dCTP (Amersham, 3000 Ci/mmol), 2 units of large fragment DNA polymerase I (Klenow; BRL) and distilled water to 50 µl. The reaction was carried

out at room temperature overnight and stopped by the addition of a solution containing 20 mM NaCl; 20 mM Tris-HCl, pH7.5; 2 mM EDTA; 0.25% SDS and 1 µM dCTP. Purification of the labelled DNA was not performed. Oligolabelling buffer was made up from the following three components: (i) two volumes of 0.018% 2-mercaptoethanol and the remaining unlabelled dNTPs (dATP, dGTP and dTTP, each 0.5 mM) in 1.25 M Tris-HCl (pH 8) and 0.125 M $MgCl_2$; (ii) five volumes of 2 M HEPES (pH 6.6); and (iii) three volumes of random sequence hexadeoxyribonucleotides (Pharmacia) at 90 OD units/ml in TE. This mixed buffer was stored at -20 °C. The extent of incorporation (typically 70-80%) of the ^{32}P labelled dCTP was assayed using Whatman DE81 paper with 0.3 M ammonium formate (pH 8) as the solvent for paper chromatography. The specific activity of the DNA labelled by this method was up to 10^9 cpm per µg of DNA when cpm were assayed using the Cerenkov channel of a scintillation counter.

2.5.11.2. End-labelling of DNA restriction fragments.

The following method was used for end-labelling isolated restriction fragments having recessed 3' termini, e.g. in the preparation of radioactive DNA size markers: to the restriction fragments were added dTTP, dGTP and dCTP to 100 µM each, 10 µCi α-[^{32}P]-dATP (Amersham, 3000 Ci/mmol) or 10 µCi α-[^{35}S]-dATP (Amersham, 600 Ci/mmol), 2 units of large fragment DNA polymerase I (Klenow; BRL) and distilled water to 50 µl. The reaction was carried out at room temperature for 30 min.

2.5.11.3. End-labelling of oligonucleotides.

Single stranded oligonucleotides were end-labelled using polynucleotide kinase and γ-[^{32}P]-ATP as follows: 0.5 µg of synthetic oligonucleotide was added to 1 µl 10 x buffer (0.5 M Tris-HCl, pH 8; 0.1 M $MgCl_2$, 50mM DTT), 60 µCi γ-[^{32}P]-ATP (Amersham, 5000 Ci/mmol) and 10 units T4 polynucleotide kinase (NBL) in a final volume of 10 µl and the mixture incubated at 37 °C for 1 h. The

labelled oligonucleotide DNA was recovered and separated from unincorporated radioactivity by two ethanol precipitations as follows: 40 µl of 2.5 M ammonium acetate and 160 µl of cold ethanol were added to the incubation mix and the DNA was precipitated at -70 °C for 15 min. The DNA was pelleted by centrifugation at 10,000 rpm for 15 min at 4 °C. The pellet was resuspended in 100 µl 0.3 M sodium acetate, 300 µl of cold ethanol was added and the DNA precipitated as before. The labelled oligonucleotides were washed in 70% ethanol and dried at room temperature. An oligonucleotide labelled in this way could be used with an appropriate unlabelled oligonucleotide in the polymerase chain reaction (see section 2.5.15.1.) to amplify a specific DNA fragment end-labelled only at the 5' end of one strand.

2.5.11.4. Labelling of a mixed population of "periodate treated" small RNAs.

"Periodate treated" tRNA (prepared by Dr. J.P. Goddard) was a mixed preparation of small placental RNAs which had been subjected to periodate oxidation and β elimination at alkaline pH after protection of the tRNAs by prior aminoacylation with all 20 amino acids as described in Goddard *et al* ., (1983) using the method of Traboni *et al* ., (1980). Only those RNAs in the mixture which were aminoacylated will have 3' OH groups capable of accepting pCp after periodate treatment. An aliquot of the periodate treated material was labelled by the ligation of [^{32}P]pCp as follows: 1 µg of periodate treated material was added to 50 µCi [^{32}P]pCp (Amersham, 3000 Ci/mmol), together with 1 µl of 10 x buffer (100 mM Tris-HCl, pH7.5; 2.5 mM DTT; 100 mM $MgCl_2$), 1 µl DMSO, 4 µl of 100 µM ATP, 10 units ofT4 RNA ligase (Anglian Biotec) and distilled water to 10 µl. The reaction was carried out at 4 °C overnight.

2.5.12. Hybridisation of blotted DNA.

2.5.12.1. Hybridisation using nylon membranes.

The following conditions were used for the hybridisation of DNA fragments (labelled by the random priming method; see section 2.5.11.1.) to denatured DNA immobilised on nylon membranes. The nylon membrane with blotted DNA was placed in a polythene bag containing 10 ml of hybridisation buffer (5 x SSPE (see section 2.2.2.), 0.5% SDS, 100 µg/ml denatured, sonicated salmon sperm DNA and 5 x Denhardt's solution (100 x Denhardt's solution contains 1% (w/v) of each of Ficoll (type 400, Pharmacia), polyvinylpyrrolidone and bovine serum albumin (Fraction V; Sigma)). The bag was sealed and the membrane pre-hybridised for 16-24 h at 65 °C with constant agitation. The appropriate DNA probe was denatured by boiling the sample for 8 min and cooling it on ice for 5 min. The probe was then added to the membrane plus prehybridisation buffer, the bag resealed and hybridisation allowed to take place for 16-24 h at 65 °C with constant agitation. The stringency of the washes used after hybridisation to remove non-specifically hybridised probe was varied as appropriate. Typically membranes were washed successively in 2 x 100 ml of 1 x SSPE for 5 min at room temperature; 2 x 100 ml of 1 x SSPE and 0.1% SDS for 30 min at room temperature; 2 x 100 ml of 1 x SSPE and 0.1% SDS at 65 °C and 2 x 100 ml of 0.5 x SSPE and 0.1% SDS at 65 °C. After washing, the membrane was wrapped with cling film and exposed to preflashed X-ray film (Fuji) using two intensifying screens (Dupont, Cronex-lightning plus) for a suitable period at -70 °C. Nylon membranes were stripped for subsequent re-use by two immersions for 10 min in distilled water plus 0.1% SDS at 100 °C.

2.5.12.2. Hybridisation using nitrocellulose membranes.

The following conditions were used for the hybridisation of mixed tRNA probes (labelled with [^{32}P]pCp; see section 2.5.11.4.) to DNA fixed to nitrocellulose

membranes. Membranes were placed in polythene bags together with 10 ml of hybridisation buffer (5 x SSC, 50% formamide) and the labelled probe added to the bag. No prehybridisation step was used, and no heat denaturation of the RNA probe is necessary. Hybridisation was allowed to take place for 16-24 h at 42 °C with constant agitation. The stringency of the washes used to remove non-specifically hybridised probe from nitrocellulose membranes was varied as appropriate. Typically membranes were washed briefly in 5 x SSC, 50% formamide then successively in 2 x 100 ml of 5 x SSC at room temperature for 30 min with gentle agitation and 2 x 100 ml of 2 x SSC at room temperature for 30 min with gentle agitation. Membranes were exposed to X-ray film as described in section 2.5.12.1.

2.5.13. Densitometric analysis of autoradiographs.

Autoradiographs obtained from experiments involving hybridisations to human genomic DNA were scanned, parallel to the direction of migration of blotted DNA, by the use of a Laser densitometer (2202 Ultrascan, LKB Bromma). Four readings were taken for each band on the autoradiograph. An example of a densitometric trace is shown in figure 4.5.B. The Laser densitometer was calibrated by Mr. J. Jardine.

2.5.14. Nucleotide sequence analysis by the Sanger chain termination method.

Nucleotide sequence analysis was performed as described by Sanger *et al* ., (1980). The DNA fragments to be sequenced were subcloned into the multiple cloning site of M13 replicative form DNA vectors (Norrander *et al* ., 1983; Yanisch-Perron *et al* ., 1985) and *E. coli* TG1 cells transformed with the recombinant DNA. Plaques containing recombinant phages were selected and single-stranded phage DNA prepared. An oligonucleotide, the M13 universal primer, was annealed to each of these DNA templates and extended in a 5' to 3'

direction by the use of DNA polymerase I large fragment (Klenow) in the presence of one of four dNTP/ ddNTP mixes. Each of these mixes contains all four dNTPs (one radioactively labelled) but only one of the four ddNTPs. The reaction conditions are such that the newly synthesised DNA strands are terminated at random lengths but at base specific positions. The resulting families of DNA strands of different lengths, each terminated by a specific dideoxynucleotide, were separated by polyacrylamide gel electrophoresis. Gels were then fixed, dried down and autoradiographed and the DNA sequence determined. Sequence data was stored using a mainframe VAX computer. Sequences were analysed using the Staden program "ANALYSEQ" (Staden, 1986). This program enables the user to search nucleic acid sequences for tRNA genes, open reading frames, repeated motifs etc. The sequences and cloverleafs shown in figures 3.6., 3.8., 3.10. and 4.10. were generated using the ANALYSEQ program.

2.5.14.1. Subcloning DNA fragments into M13 vectors and transformation of *E. coli* TG1.

Restriction digests and ligations of vector and insert DNAs were performed as described in sections 2.5.6. and 2.5.8. respectively. Half (5 µl) of each ligation mix was added to 300 µl of competent *E. coli* TG1 cells (prepared as described in section 2.5.9.) and the mixture left on ice for 2-3 h. The cells were then heat shocked at 42 °C for 3 min and returned to ice. During this time the following were added to 3 ml of molten H-top agar (see section 2.2.1.) kept at 45 °C : 35 µl IPTG (100 mM), 35 µl X-gal (2% in dimethylformamide) and 200 µl fresh TG1 cells. Immediately after mixing, the molten agar was added to the transformation mix and spread onto a prewarmed H-plate (see section 2.2.1.). The plate was allowed to set and incubated at 37 °C overnight. Transformed cells harbouring recombinant M13 phage appeared as opaque colourless plaques on a lawn of uninfected cells, while those containing religated or uncut vectors appeared as blue plaques.

2.5.14.2. Preparation of single-stranded DNA templates.

A single white plaque was inoculated into 1.5 ml 2 x YT (see section 2.2.1.) containing 0.01 volume of an overnight culture of *E. coli* TG1. This culture was vigorously agitated for 4.5 h at 37 °C and the cells sedimented by centrifugation at 10,000 rpm for 5 min. The supernatant containing extruded phage was transferred to a fresh tube and recentrifuged to remove residual cells. The second supernatant was added to 200 µl of a solution of 20% (w/v) polyethylene glycol 6000 and 2.5 M NaCl and the mixture left standing at room temperature for 15 min. The precipitated phage was then collected by centrifugation at 10,000 rpm for 5 min and the supernatant aspirated using a drawn out Pasteur pipette to ensure removal of all traces of polyethylene glycol. The phage was resuspended in 100 µl of TE, extracted once with phenol and once with chloroform, ethanol precipitated, dried, redissolved in 25 µl TE and stored at -20 °C.

2.5.14.3. Sequencing of single-stranded DNA templates.

Sequencing was carried out using the protocols supplied by Amersham plc in the M13 Cloning and Sequencing Handbook.

2.5.14.3.1. Annealing and sequencing reactions.

The primer used (5'd[GTAAAACGACGGCCAGT] 3') was the 17-mer M13 universal primer with a sequence complementary to that of a region adjacent to the multiple cloning site of M13. To anneal the primer to the template, single-stranded template DNA (approximately 1 µg) was mixed with 20 ng of the primer in a final volume of 10 µl in Klenow reaction buffer (10 mM Tris-HCl, pH 8.5; 5 mM $MgCl_2$), and incubated at 60 °C for 90 min. Subsequent steps were carried out at room temperature. After annealing, 15 µCi α-[^{35}S]-dATP (Amersham, 600 Ci/mmol) and two units of Klenow fragment were added to the annealed

primer/template. A 2.5 µl aliquot of the mixture was placed into each of four tubes marked A, G, C and T in a rack for an Eppendorf 5413 microcentrifuge. A 2 µl aliquot of the relevant dNTP/ddNTP mix (see below) was placed separately into each tube. A brief spin mixed the contents of the tube and started the reaction. After 20 min, 2 µl of chase mix (a uniform 0.5 mM solution of all four dNTPs) was placed into each tube. After a second brief spin the chase was allowed to procede for 15 min. The chase reaction was stopped by the addition of 4 µl of formamide dye (0.03% xylene cyanol FF, 0.03% bromophenol blue and 20 mM EDTA in deionised formamide). The dNTP/ddNTP mixes used were varied according to the nature of the DNA template and the length of sequencing run needed.The basic mixes for sequencing with α-[^{35}S]-dATP were made up as shown below:

Deoxy NTP mixes (A^o, C^o, G^o, T^o):

	A^o	C^o	G^o	T^o
0.5 mM dCTP	20 µl	1 µl	20 µl	20 µl
0.5 mM dGTP	20 µl	20 µl	1 µl	20 µl
0.5 mM dTTP	20 µl	20 µl	20 µl	1 µl
TE, pH 8	20 µl	20 µl	20 µl	20 µl

Dideoxy NTP working solutions were: 0.1 mM ddATP, 0.1 mM ddCTP, 0.3 mM ddGTP and 0.5 mM ddTTP.

The standard dNTP/ddNTP mix was made up of equal volumes of a dNTP mix (A^o, G^o, C^o or T^o) and the corresponding ddNTP working solution. For sequencing DNA rich in G+C, the dGTP was replaced by the nucleotide analogue 7-deaza-dGTP at the same concentration. The use of 7-deaza-dGTP helps to prevent the "compressions" in a sequence caused by secondary structure formation in the template DNA, particularly in G+C rich regions (Martin, 1987). The ratio of [dNTP]:[ddNTP] was increased for longer sequencing runs. Reducing the concentration of ddNTP causes the average length of DNA chains synthesised to increase by making termination with a ddNTP less likely.

2.5.14.3.2. Gel electrophoresis and autoradiography.

Polyacrylamide-urea denaturing gels were prepared as follows: a 40% stock solution of acrylamide was prepared (see section 2.5.7.1.) of which 7.5 ml was added to 21 g of urea (BRL ultra PURE) and the mixture made up to 50 ml with 1 x TBE buffer (see section 2.2.2.). To the mixture 300 µl of fresh 10% ammonium persulphate and 50 µl of TEMED were added and the gel (40 x 20 x 0.04 cm) poured between two clean glass plates, using a sharkstooth comb (BRL). The notched gel plate was treated with dimethyldichlorosilane to prevent the gel adhering to both plates. The gel was left for at least one hour to polymerise fully then transferred to the sequencing gel apparatus (61 x 53 Vertical slab unit 400; Shandon, Southern), the reservoirs of which were filled with 1 x TBE buffer. The sequencing reaction samples were boiled for at least 3 min then loaded immediately onto the gel. Electrophoresis was carried out at 30 mA and 40 W usually until the bromophenol blue reached the bottom of the gel. After electrophoresis and removal of the notched gel plate the gel was fixed by soaking in a two litre bath of 10% (v/v) acetic acid; 10% (v/v) methanol for 15-20 min to remove the urea. The gel was drained for a few minutes, transferred onto a sheet of Whatman 3 MM paper, covered with cling film and dried under vacuum on a gel drier (model 1125B Bio-Rad) for 1 h at 80 ^{o}C. After drying the gel was exposed directly onto Fuji X-ray film overnight at room temperature. Longer exposures were used as necessary. Following autoradiography it was usually possible to read over 200 nucleotides from one loading. Two separate loadings were necessary to maximise the length of sequence that could be read: the first loading was subjected to electrophoresis for 3-4 h and the second loading for 1.5-2 h.

2.5.15. The polymerase chain reaction (PCR).

The polymerase chain reaction is a powerful method for the *in vitro* amplification of specific DNA fragments (for review see White *et al*., 1989). The methods described in sections 2.5.15.1. and 2.5.15.2. are based on protocols given in Innes *et al*., (1990).

2.5.15.1. Basic PCR methodology.

Standard PCR amplification reactions were set up in a total volume of 100 µl. The reaction mix contained: 10 µl of Promega 10 x Taq DNA polymerase buffer (500 mM KCl, 100mM Tris-HCl, pH 9, 15 mM $MgCl_2$, 0.1% gelatin (w/v), 1% Triton X-100), 0.2 mM dATP/dGTP/dCTP/dTTP, 2.5 units Taq DNA polymerase (Promega), 0.5 µg of each of the two oligonucleotide primers used and a suitable amount of target DNA. The components were added in the order shown. The amount of target used depended on the complexity of the DNA containing the DNA fragment to be amplified. Typical amounts used were 2 ng of recombinant plasmid DNA, 5 ng of recombinant λ DNA and 0.2 µg of human genomic DNA. Amplifications were carried out using a programmable DNA thermal cycler (Perkin Elmer Cetus). The exact conditions used for amplification varied with the nature and length of the target sequence. For G+C rich targets the 0.2 mM dGTP in the mix was replaced with 0.2 mM 7-deaza-dGTP. The stages involved in the PCR were an initial denaturation of the target DNA at 94 °C, followed by repeated cycles of denaturation (94 °C), annealing of primers (50-60 °C), new DNA synthesis (72 °C) and a final single extended period at 72 °C to ensure that the majority of amplified products were of full length. A typical set of conditions used was: 94 °C for 3 min, 30 x [94 °C, 1min; 55 °C, 30 s; 72 °C, 2 min], 72 °C for 5 min. Finally, samples were cooled to 4 °C to stop the reaction. PCR amplified DNA fragments were analysed by agarose or polyacrylamide gel electrophoresis.

2.5.15.2. Production of single-stranded DNA by asymmetric PCR.

The PCR can be used to produce single-stranded DNA which is a suitable template for sequencing by the Sanger chain termination method (see section 2.5.14.). The method used (McCabe, 1990) is termed asymmetric PCR since it involves the use of an unequal concentration of the two amplification primers. This modification of the PCR protocol leads to an initial amplification of double stranded

product until the lower concentration primer becomes depleted, following which an excess of one of the two strands is generated, depending on which of the amplification primers was limited. This single-stranded DNA accumulates linearly and is complementary to the limiting primer which can subsequently be used to prime sequencing reactions. The complementary strand can be generated for sequencing by reversing the relative concentrations of the two primers. Asymmetric PCR was used in this study to sequence genomic copies of tRNA genes. The reaction mix for asymmetric PCR was prepared essentially as described in section 2.5.15.1. using 0.2 µg of target human genomic DNA. The ratio of limiting : non-limiting primer used was 20 : 1 (1 µg : 50 ng). The optimum conditions found for the amplification reaction were 94 °C for 3 min, 45 x [94 °C, 1min; 55 °C, 1 min; 72 °C, 20 s], 72 °C for 3 min. After the addition of 2.5 units of fresh Taq DNA polymerase a further 10 cycles were performed. The amplified DNA fragments were precipitated by the addition of ammonium acetate to 2 M and isopropanol to 50%, pelleted by centrifugation at 10,000 rpm for 10 min at 4 °C, washed with 70% ethanol and dried. Analysis by agarose gel electrophoresis confirmed that one discrete band was predominantly amplified, although some background bands were visible. The single-stranded DNA prepared in this way proved to be suitable for sequence analysis without further purification.

2.5.16. *In vitro* transcription of tRNA genes.

2.5.16.1. HeLa S3 cell extract.

HeLa cell nuclear extract was prepared using an adaptation of the method of Weil *et al* ., (1979). HeLa S3 cells (Gibco) were grown in suspension in 700 ml stirring flasks (Techne, Cambridge) in the Glasgow modification of Eagle's minimal essential medium (see Adams, 1980; provided by Gibco), supplemented with 10% calf serum, 1% glutamine, 1% non-essential amino acids and 1 M NaOH to give a pH of approximately 7.5. The cells were monitored daily by microscopic examination. Cell culture was carried out by Dr. R.L.P. Adams, Mrs. A. Rinaldi

and Mrs. H. Lindsay. Cells from three litres of culture (at a density of approximately 5 x 10^5 cells/ml) were harvested and washed with cold PBS mix (see section 2.2.2.). The packed cell volume was measured after centrifugation at 1,000 rpm in a Beckman TJ-6 centrifuge for 5 min at 4 °C. The cells were then washed with hypotonic buffer (10 mM HEPES, pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM DTT) and the cell pellet was resuspended in twice the packed cell volume of hypotonic buffer. The cells were allowed to swell for 10 min on ice and then disrupted in a Kontes all glass Dounce homogeniser by 18-20 strokes of the B pestle. The extent of cell lysis was generally 90-95% as monitored by light microscopy, while nuclear lysis was minimal. One tenth volume of a solution containing 0.3 M HEPES, pH 7.9, 1.4 M KCl and 30 mM $MgCl_2$ was added and the lysate was centifuged at 40,000 rpm in a Ti50 rotor of a Beckman ultracentrifuge for 1 h at 4 °C. The supernatant was placed in a fresh tube and glycerol added to 20% (v/v). The cell free extract was stored in 400 µl aliquots at -70 °C. All of the transcription assays mentioned in this study were performed using aliquots from a single HeLa cell extract preparation. The protein concentration of this crude extract preparation was determined as 1 mg/ml by Ms. C. Houlston using the method of Bradford (1976).

2.5.16.2. *In vitro* transcription assay.

The method used for the transcription assays described in this study was optimised in our laboratory by Dr. E. Gonos. The protocol was initially based on a consensus of the methods of Laski *et al* ., (1983), Makowski *et al* ., (1983), Murphy and Baralle (1983), Schaak *et al* ., (1984) and Segall *et al* ., (1980). Transcription assays were normally performed in duplicate. A typical transcription reaction mixture (20 µl) contained final concentrations of 14 mM HEPES-KOH, pH 7.9, 80 mM KCl, 20 mM NaCl, 3.5 mM $MgCl_2$, 10% (v/v) glycerol, 0.3 mM DTT, 600 µM unlabelled ATP, CTP and GTP, 2.5 µCi of α-[^{32}P]-UTP (Amersham, 400 Ci/mmol), 25 µM unlabelled UTP, 10 mM creatine phosphate, 10 µl HeLa cell free extract, the desired amount of supercoiled recombinant DNA and pUC19 DNA carrier to 100-200 ng of total DNA in the assay. (It was found that the addition of pUC19 carrier DNA (100-200 ng) to transcription assays increased the

efficiency of transcription of recombinants several fold, without itself giving rise to any transcripts. Experiments also showed that transcripts could be produced from isolated restriction fragments containing tRNA genes used in *in vitro* transcription assays, but with a far lower efficiency than that found for supercoiled plasmid templates.) In preliminary assays α-amanatin (final concentration 10 μg/ml) was added to the transcription reaction mixture to eliminate the possibility that RNA polymerase II was responsible for the transcripts obtained (see Jacob, 1973). The mixture was incubated at 30 °C for 1 h (except in the case of time course assays) and the reaction stopped by the addition of 20 μl of a solution containing 1% SDS, proteinase K (2 mg/ml) and *E. coli* crude tRNA (1 mg/ml). The reaction mixture was maintained at 37 °C for 20 min and then an equal volume of 1 M ammonium acetate was added. The mixture was extracted once with phenol/chloroform and the aqueous layer was tranferred to a fresh Eppendorf tube. An equal volume of 0.5 M ammonium acetate was added to the phenol/chloroform and after the back extraction the second aqueous layer was added to the first. Cold ethanol was added to 70% and the RNA precipitated at -70 °C for 30 min and pelleted by centrifugation at 10,000 rpm for 15 min at 4 °C. The pellet was dried briefly, resuspended in 12 μl TE and mixed with 10 μl formamide dye mix (see section 2.5.14.3.1.). The samples were then loaded onto a 10% polyacrylamide-4M urea vertical gel (prepared essentially as described in section 2.5.14.3.2.) and electrophoresed at 250 volts for 1.5-2 h using a BRL model V161 gel apparatus. After electrophoresis one gel plate was removed, the gel covered with cling film and exposed to X-ray film (Fuji) in the presence of an intensifying screen at -70 °C. (Alternatively, samples were loaded onto a sequencing gel (section 2.5.14.3.2.) for better resolution of transcripts. The size of transcripts could be estimated by comparison to DNA fragments produced by a sequencing reaction and electrophoresed alongside the assay samples). The basic assay method shown above was adapted as necessary for the experiments described in Chapter five.

2.5.16.3. Analysis of transcripts by the use of a scintillation counter.

Soon after autoradiography, autoradiographs were aligned to their corresponding

gels by the help of radioactive markers and the regions of the gel containing the labelled transcripts were excised and transferred to scintillation vials. Other regions of the gel with the same dimensions (where no transcripts were detected) were also excised for use in calculating background. (The same basic approach was used in quantifying transcripts separated on sequencing gels which were subsequently dried down onto Whatman 3 MM paper). Samples were Cerenkov counted for 5 min in a Beckman LS6800 scintillation counter, and the background value subtracted from each sample. An average value was established for each set of duplicates, any decay factor calculated and finally the value of each transcript expressed in counts per minute (cpm).

2.5.16.4. Calculation of tRNA products.

The tRNA products derived from *in vitro* transcription assays were calculated according to a formula set out by Gonos (1989). The following example is for a transcription assay using a pUC19 recombinant pGLY1 (see table 5.1.) which contains a $tRNA^{Gly}$ gene:

In one hour, 2 ng of pGLY1 produced 1300 cpm of transcripts. The efficiency of Cerenkov counting was found to be 35-40%. Therefore the transcripts produced per ng of DNA in one hour = 1733 dpm (A). In every 20 µl assay, 5 µCi of α-[^{32}P]-UTP (400 Ci/mmol) were used, which corresponds to 0.6 µM. In addition, 25 µM of unlabelled UTP were also used; therefore the total UTP concentration is 25.6 µM which corresponds to 512 pmol (B) in 20 µl .The 5 µCi of α-[^{32}P]-UTP used corresponds to 1.1×10^7 dpm (C). It follows that [(A/C) x B] = 80.7 fmol UTP were incorporated into RNA (D). The $tRNA^{Gly}$ primary transcript contains 20 uridines (E) therefore (D/E) the transcripts produced by 1 ng pGLY1 in 1 hour are 4.04 fmol tRNA (F). Since 1 ng pGLY1 corresponds to 0.45 fmol (G) (pGLY1 molecular weight = 2.2×10^6) then (F/G) pGLY1 produces approximately 9 transcripts/gene/hour.

2.5.17. Gel shift assays.

Oligonucleotide primers complementary to either end of a 175 bp fragment from the 5' flank of a $tRNA^{Gly}$ gene from λHt363 (see section 5.6.3.) were end labelled as described in section 2.5.11.3. Two polymerase chain reactions were performed as described in section 2.5.15.1., using in each reaction one labelled and one unlabelled oligonucleotide primer, both using as target DNA a recombinant plasmid (pGLY1, see table 5.1.) containing the 175 bp fragment in its insert. The products of the two reactions were the amplified 175 bp fragment, in one case labelled at the 5' position of one strand and in the converse reaction at the 5' end of the other. Aliquots of these end labelled fragments were mixed with a HeLa cell free extract (containing 1.5-7.5 µg total protein) in a total volume of 20 µl and incubated for 30 min at 30 °C. The same ionic conditions were employed as those used for *in vitro* transcription assays (see section 2.5.16.2.). The samples were added to 10 µl of loading buffer (see section 2.5.7.1.) and loaded onto a 10% polyacrylamide gel. The gels were run at 10 V/cm for 1-2 h using a BRL model V161 vertical gel apparatus. After electrophoresis gels were dried under vacuum on a gel drier (model 1125B Bio-Rad) at 80 °C for one h and exposed to X-ray film (Fuji) without intensifying screens.

2.5.18. DNase I protection assays.

The method used was adapted from that described by Bothwell (1990). End labelled 175 bp fragments were incubated in a HeLa cell free extract as described in section 2.5.17. At 30 min 20 µl of 0.01 M $MgCl_2$ and 1 µl of DNase I solution (varying between 0.05 and 2 µg per µl) were added. After 1 min at room temperature the DNase I digestion was stopped by the addition of 1 µl of 0.5 M EDTA and the samples placed on ice. The samples were then added to 10 µl of loading buffer and electrophoresed on a 10% polyacrylamide gel as described in section 2.5.17. Gels were autoradiographed as described in section 2.5.16.2. The optimum DNase I amount used (the highest to give a gel shift pattern the same as that obtained in untreated samples) was found to be between 1 and 2 µg.

Autoradiographs were aligned to the gels with the help of radioactive markers and the regions of the gel containing the free and protein-bound DNase I treated DNA fragments excised. The DNA was eluted in the following way. The gel fragments were crushed and placed in Eppendorf tubes with 0.5 ml of elution buffer (0.5 M ammonium acetate, 0.1 M EDTA, 0.1% SDS, 10 μg/ml proteinase K and 5 μg/ml carrier tRNA). Samples were incubated at 37 °C overnight. The eluted DNA was extracted once with phenol/chloroform and once with phenol, ethanol precipitated by the addition of two volumes of cold ethanol and pelleted by centrifugation at 10,000 rpm for 15 min at 4 °C. The pellets were washed with 70% ethanol and dried briefly. The samples were electrophoresed on a sequencing gel as described in section 2.5.14.3.2. and autoradiographs produced by placing X-ray film (Fuji) in contact with the dried gel at -70 °C with intensifying screens.

CHAPTER 3

RESTRICTION MAPPING AND SEQUENCING OF λHt363.

3.1. Background and aims.

The recombinant phage λHt363 was selected from a human genomic library, cloned in λ Charon 4A (Lawn *et al* ., 1978) using a mixed tRNA probe labelled by ligation of [^{32}P]pCp (Goddard *et al* ., 1983). The recombinant had been partially characterised (McLaren and Goddard, 1986; J.P. Goddard, A.McLaren, T.Carr and D.Livingstone, unpublished results). A 4.2 kb fragment had been subcloned into the plasmid vector pAT153 (Twigg and Sherratt, 1980). Restriction fragments within a 1.65 kilobase (kb) region on this recombinant plasmid were shown to hybridise to a mixed tRNA probe. Three restriction fragments from within this 1.65 kb region had been sequenced. The sequences of two *Eco* RI-*Hin* d III restriction fragments of 246 and 228 base pair (bp) had been determined on both strands. The 248 bp fragment contains a gene coding for tRNA^{Leu} (McLaren and Goddard, 1986) and the 228 bp fragment a gene coding for tRNA^{Gln}. A 439 bp *Bam* HI- *Kpn* I restriction fragment had been sequenced on one strand only and found to contain a gene coding for tRNA^{Lys}. These features of the insert of λHt363 resemble those of a human tRNA gene cluster (cloned from the same genomic library) which has previously been reported on a recombinant termed λHt4 (Roy *et al* ., 1982, Buckland *et al* ., 1983). This cluster contains the same three tRNA genes, coding for tRNA^{Lys}, tRNA^{Gln} and tRNA^{Leu}, also within a 1.65 kb region. The published sequence from λHt4 (Roy *et al* ., 1982) appeared to be almost identical to that so far found on λHt363.

In addition to the 1.65 kb region mentioned above, an 800 bp *Eco* RI restriction fragment from λHt363 also hybridised to the mixed tRNA probe. This fragment had been partially sequenced and shown to contain the 3' end of a tRNA^{Gly} gene, consisting of 51 bases starting from one of the *Eco* RI sites and including a putative

termination signal.

The aim of this part of the project was to complete the restriction map of the insert of λHt363. Once the map was finalised, the position of the tRNA genes within the insert would be established, and the sequencing of the genes and their flanking regions completed. These sequences and the restriction map could then be compared with the apparently similar human tRNA gene cluster published (Roy *et al* .,1982; Buckland *et al* ., 1983).

3.2. Restriction mapping of λHt363.

3.2.1. Construction of a preliminary restriction map of λHt363.

Restriction endonuclease digests were performed as described in section 2.5.6. A preliminary restriction map for the insert of λHt363 was constructed by the estimation of the fragment sizes produced by digestion with a series of restriction enzymes, singly and in combination. The enzymes used initially were the six-base cutters *Bam* HI, *Bgl* II, *Eco* RI, *Hin* d III and *Kpn* I. A unique site for the eight-base cutter *Not* I was also discovered within the insert. Restriction digests were analysed by agarose gel electrophoresis (see section 2.5.7.1.). Figure 3.1. shows the pattern of fragments observed on agarose gel electrophoresis of digests using each of these enzymes and table 3.1. shows the fragment sizes obtained. Restriction data for the enzymes *Apa* I and *Sma* I is also presented in figure 3.1 and table 3.1. These digests were performed at a later stage in the study when it became apparent that a large region of the insert of λHt363 had a high (G+C) content. Since the recognition sites for *Apa* I and *Sma* I consist only of C or G residues these enzymes were useful in refining the preliminary restriction map. The total insert size of the recombinant was estimated at this stage as 12.5-13.5 kb.

3.2.2. Ordering of restriction fragments from λHt363.

Having ascertained the numbers and sizes of the restriction fragments of λHt363,

Figure 3.1. Restriction digests of λHt363 DNA.

This photograph shows an ethidium bromide stained 0.8% agarose gel where the following λHt363 restriction digests were electrophoresed: *Apa* I (A), *Bam* HI (B), *Bgl* II (Bg), *Eco* RI (E), *Hin* d III (H), *Kpn* I (K), *Not* I (N) and *Sma* I (S). The marker DNA fragment sizes are indicated (in kb). M1 = M13mp18 DNA/*Hin* f I, M2 = λ DNA/*Hin* d III.

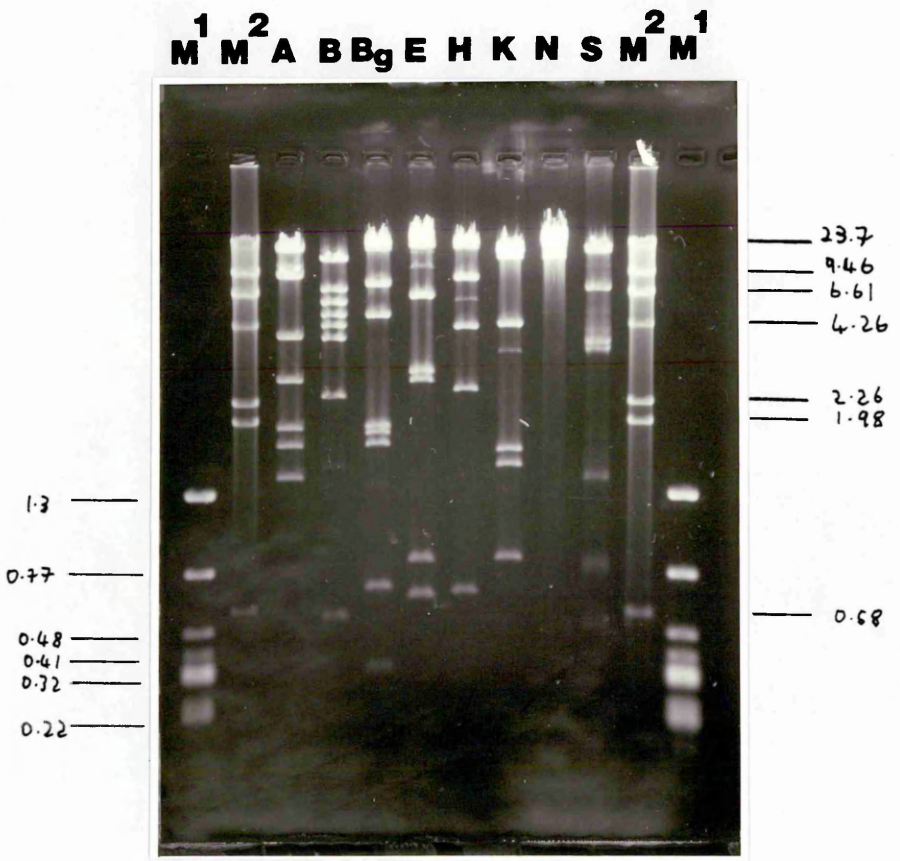


Table 3.1.

Restriction fragment profile of λHt363.

Apa I	*Bam* HI	*Bgl* II	*Eco* RI	*Hin* d III	*Kpn* I	*Not* I	*Sma* I
12.1§	15§	24.4§	19.9*	22.1§	17.4*	29.5§	20.2§
10.8§	6§	7.8	11*	8.9§	14.8§	14.5§	7.3*
10*	5.6*	4.9*	6.2	5.7*	4.2		4§
3.7	4.8	2.2*	2.9	4.2	3.5*		3.7
2.6	4.2	1.9§	2.6	2.4	1.75§		3.4
1.85	3.9*	1.3*	0.8	0.6	1.5*		1.4
1.6	2.4	0.65*	0.6		0.84		0.85
1.35	1.5*	0.47*					0.6
0.18	0.5*	0.36*					0.6
	0.12	0.06*					0.44
							0.43
							0.35
							0.31
							0.28
							0.24

λHt363 total length =44 kb. Insert length =13 kb.

All sizes are expressed in kb. * denotes fragments wholly derived from the arms of λ Charon 4A and § fragments derived in part from the λ arms.

these were then ordered using the end-labelling of isolated restriction fragments generated by single or double restriction enzyme digests (see section 2.5.11.2.). Either before or after end-labelling, isolated restriction fragments were completely or partially digested with a second restriction enzyme, and the resulting fragments separated by agarose gel electrophoresis. The gel was then dried down and autoradiographed. Figure 3.2. shows the autoradiograph obtained from one such experiment. As mentioned in section 3.2.1., the restriction enzyme *Not* I was found to have a unique site in the recombinant λHt363, located in the insert. Digestion with *Not* I thus generated two fragments which were separated by agarose gel electrophoresis. End-labelling of these separated fragments, followed by further digests with a second enzyme or enzymes, was particularly useful in ordering restriction sites.

3.2.3. Refining the restriction map of λHt363.

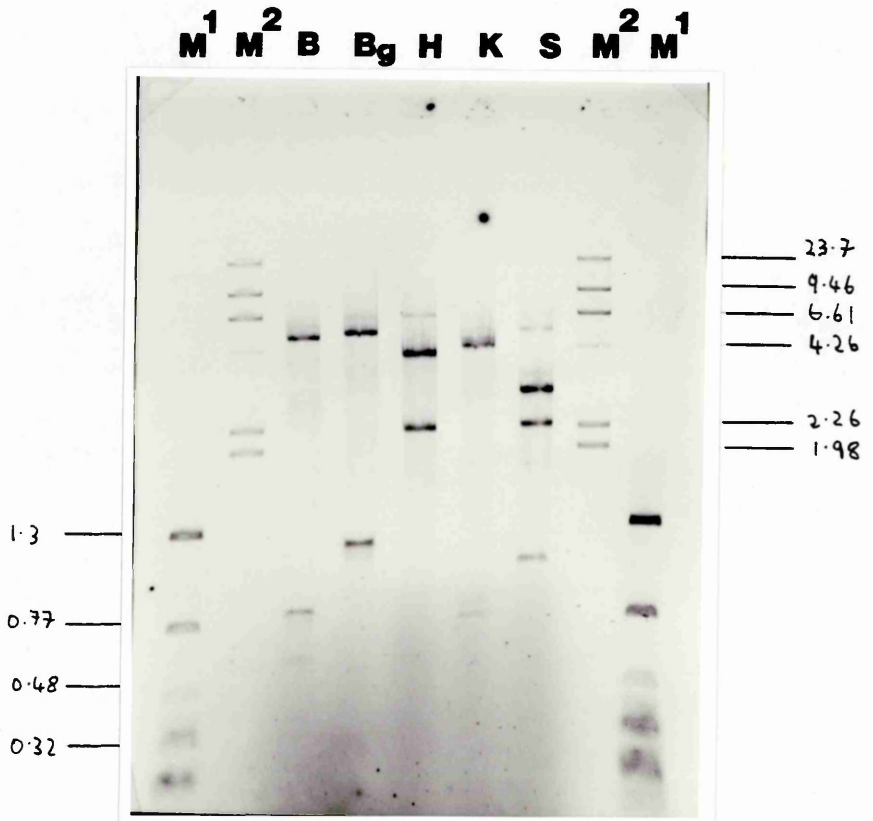
The construction of a preliminary restriction map of λHt363 allowed the selection of a set of restriction fragments (five in total) spanning the whole insert. These were subcloned into pUC19 (see section 2.5.8.). Restriction digests of these subcloned fragments were used to remove any remaining ambiguities in the restriction map, which is shown in figure 3.3. The insert size of λHt363 was estimated as 13kb ± 300 bp. The positions of the five tRNA genes eventually found in the cluster are indicated in figure 3.3. (Earlier unpublished hybridisation data (J. P.Goddard, A.McLaren and T.Carr) indicated the restriction fragments which hybridised to a mixed tRNA probe). Figure 3.3. also shows the regions of λHt363 for which sequence data has been obtained (see section 3.3.) and the location of the pUC19 subclones used in the construction of the restriction map.

3.3 Sequencing of the tRNA gene containing regions of λHt363.

Having constructed the restriction map of λHt363, and located the tRNA genes, it became possible to isolate the further subclones necessary for completion of the

Figure 3.2. Analysis of end-labelled restriction digests of an isolated 6.2 kb *Eco* RI restriction fragment from λHt363.

A 6.2 kb *Eco* RI fragment from λHt363 was isolated by agarose gel electrophoresis and digested with the restriction enzymes indicated below (*Bam* HI (B), *Bgl* II (Bg), *Hin* d III (H), *Kpn* I (K) and *Sma* I (S)). The resulting fragments were end-labelled with [^{32}P]dATP by filling in the recessed ends left by restriction digests with the Klenow fragment of DNA polymerase I. The labelled fragments were separated by electrophoresis on a 0.8% agarose gel which was dried down and autoradiographed overnight. Marker DNAs were end-labelled by the same method with [^{35}S]dATP. Fragment sizes are indicated (in kb). M1 = M13mp18 DNA/*Hin* fI, M2 = λ DNA/*Hin* d III.



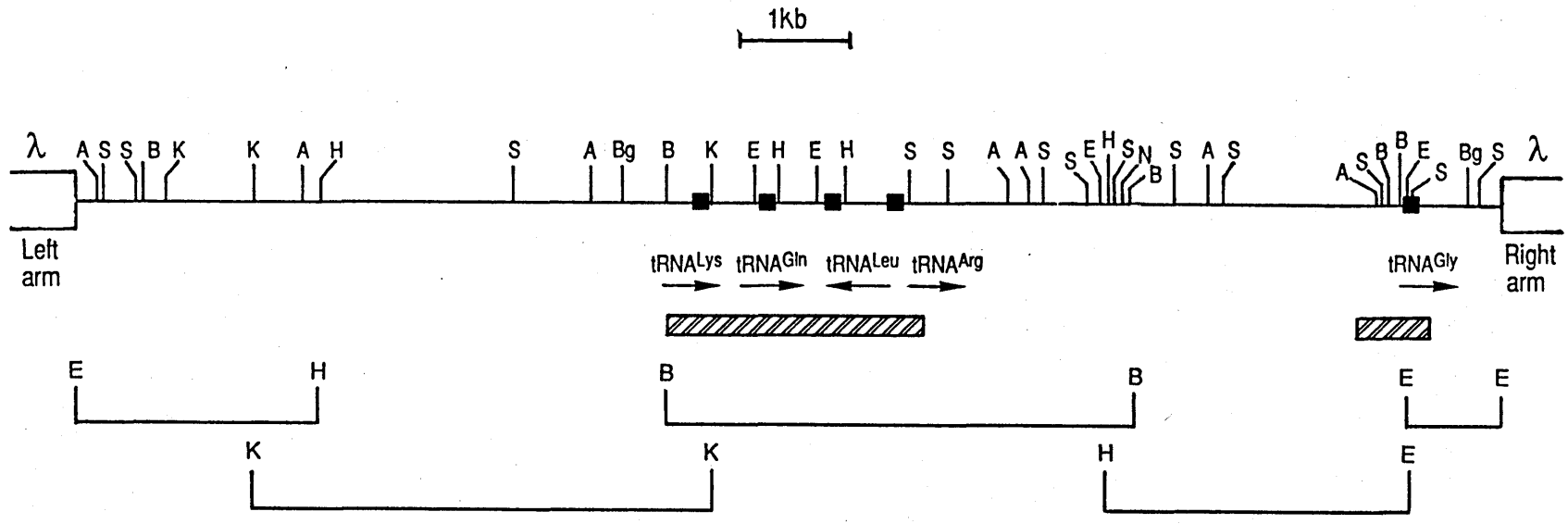


Figure 3.3. Restriction map of the insert of λHt363.

The sites for the restriction enzymes *Apa* I (A), *Bam* HI (B), *Bgl* II (Bg), *Eco* RI (E), *Hin* d III (H), *Kpn* I (K), *Not* I (N) and *Sma* I (S) within the insert of λHt363 are shown, together with the positions (■), orientations and identities of the tRNA genes found on λHt363. Hatched boxes indicate the extent of the nucleotide sequence determined. Five of the overlapping restriction fragments spanning the insert of λHt363 which were subcloned into pUC19 are shown.

sequencing of the genes and their flanking regions. Sequencing reactions were performed using the chain termination method (Sanger *et al*., 1980) as described in section 2.5.14.3. Restriction fragments were subcloned into the M13 vectors mp10, mp11, mp18 or mp19 (Norrander *et al*., 1983; Yanisch-perron *et al*., 1985). Subcloning and the preparation of single-stranded DNA were performed as described in sections 2.5.14.1 and 2.5.14.2.

3.3.1. Sequencing of the $tRNA^{Gly}$ gene region.

Prior to the start of this project, the partial sequencing of an 800 bp *Eco* RI fragment had revealed an incomplete $tRNA^{Gly}$ gene sequence (D.Livingstone, unpublished results). The remainder of the sequence of the $tRNA^{Gly}$ gene and its 5' and 3' flanks was determined using subclones of a 713 bp *Pvu* II-*Stu* I fragment. Figure 3.4. shows an autoradiograph of a sequencing gel using single-stranded template DNA prepared from subclone M1 (see figure 3.5. and table 3.2.) which contains the $tRNA^{Gly}$ gene.The DNA was sequenced on both strands, reading through all restriction sites.The sequencing strategy is summarised in figure 3.5. and the M13 subclones used are listed in table 3.2. The sequence is presented in figure 3.6., which also shows the position and orientation of the $tRNA^{Gly}$ gene and the predicted cloverleaf secondary structure of the RNA-like strand of the gene.The sequence of the gene and its immediate flanks proved to be identical to a 397 bp sequence published during the course of this study (Doran *et al*., 1988). Both of these recombinants were cloned from the same genomic library, but the restriction maps of λHt363 and the recombinant cloned by Doran *et al*. (λ Ht9) differ sufficiently to show that they represent different parts of the genome (see section 3.4. and figure 3.11.). The $tRNA^{Gly}_{GCC}$ gene from λHt363 has an identical sequence to that of a human placental $tRNA^{Gly}_{GCC}$ (Gupta *et al*., 1979).

3.3.2. Sequencing of a 1651 bp region containing three tRNA genes.

A 1.65 kb region containing tRNA genes coding for $tRNA^{Lys}$, $tRNA^{Gln}$ and

Figure 3.4. Autoradiograph of a sequencing gel.

This figure shows the autoradiograph of a sequencing gel for the M13mp18 subclone M1 (see table 3.2.). The 237 bp *Bam* HI-*Stu* I insert of this subclone contains a $tRNA^{Gly}$ gene. The sequence of the RNA-like strand of the gene is indicated, representing residues 503 to 573 inclusive of the sequence shown in figure 3.6.

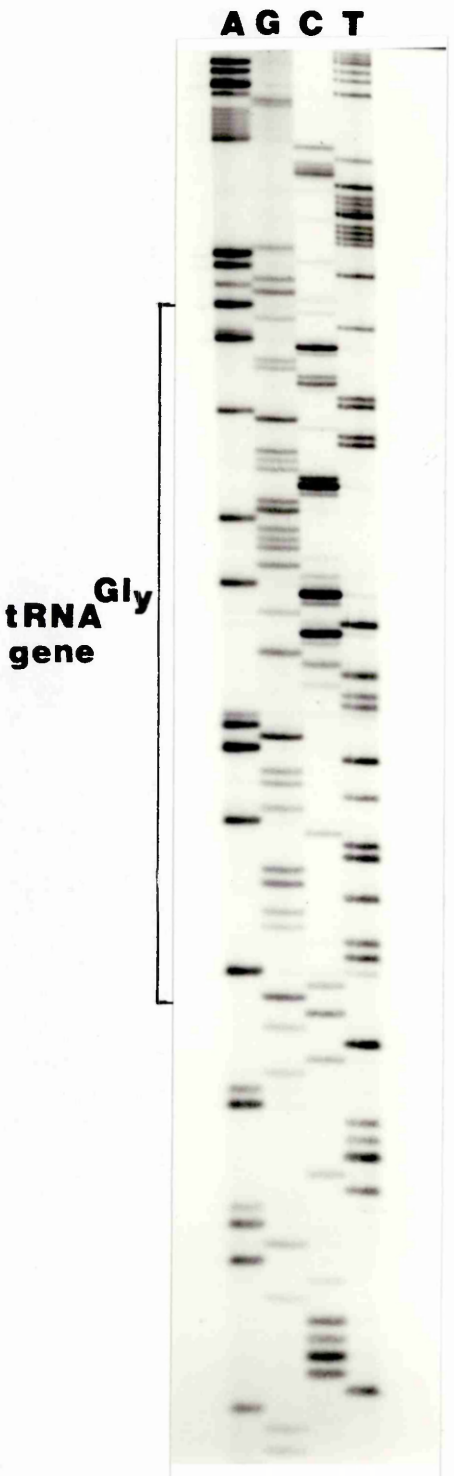


Figure 3.5. Strategy used for sequencing a 713 bp region from λHt363 containing a $tRNA^{Gly}$ gene.

This figure summarises the restriction sites used (*Bam* HI (B), *Eco* RI (E), *Pvu* II (P), *Sma* I (S) and *Stu* I (St)) to generate M13 subclones for sequencing a 713 bp region of λHt363. The arrows indicate direction and extent of sequence determined from each subclone (see also table 3.2.).

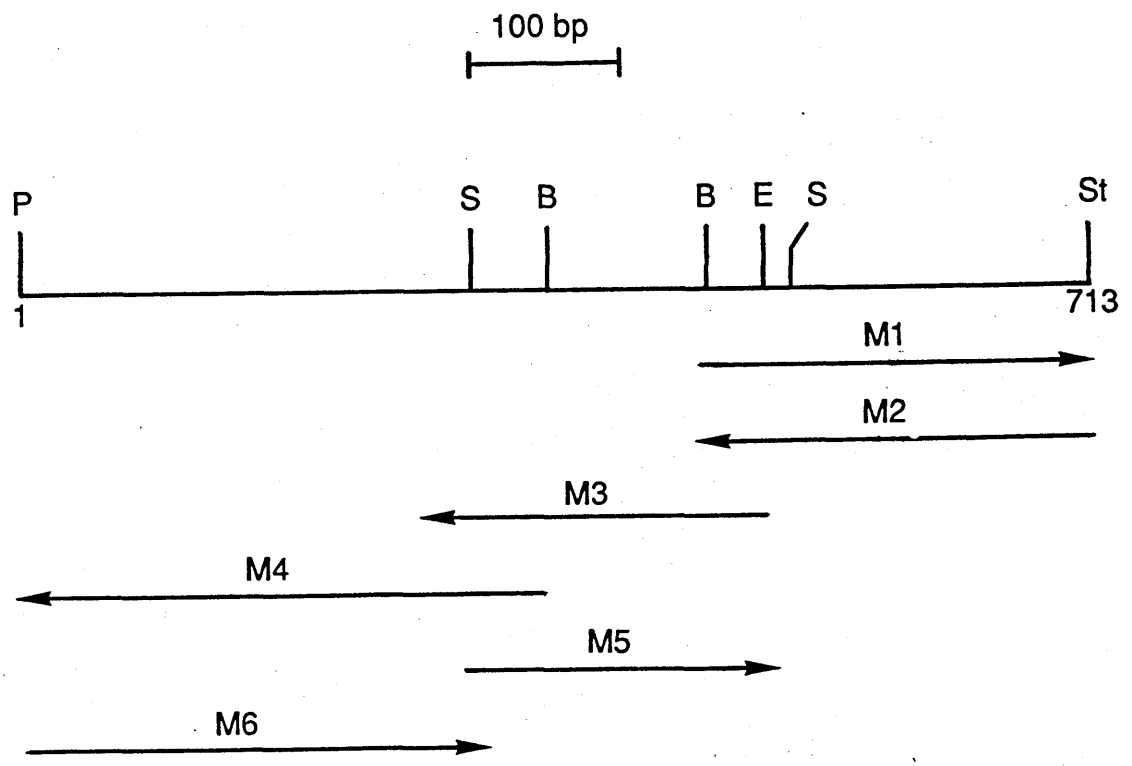


Table 3.2.

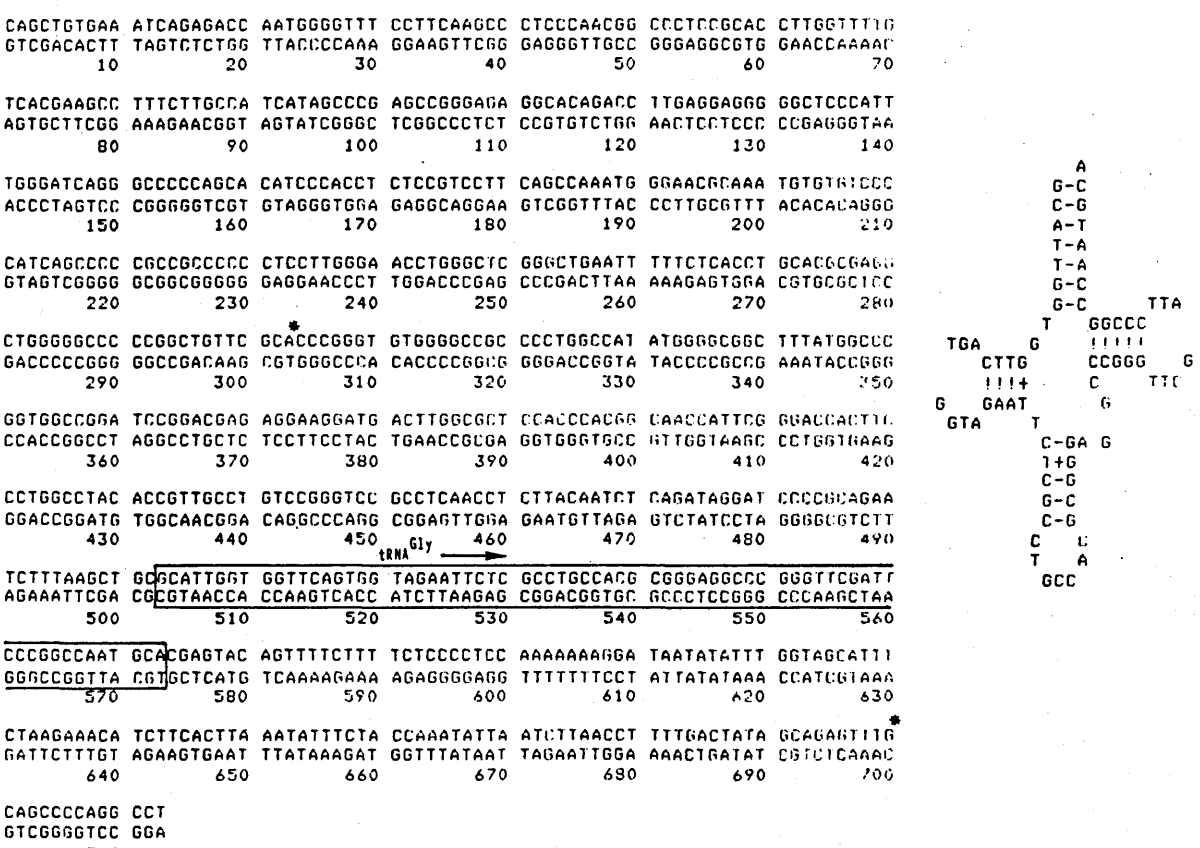
Subclones used in sequencing a 713 bp region containing the $tRNA^{Gly}$ gene from λHt363.

Subclone	Insert		Vector	Sequenced region
M1	*Bam* HI-*Stu* I	477-713	M13mp18	477-713
M2	*Stu* I-*Bam* HI	713-477	M13mp19	713-477
M3	*Eco* RI-*Pvu* II	528-1	M13mp19	528-257
M4	*Bam* HI-*Pvu* II	363-1	M13mp18	363-1
M5	*Sma* I-*Eco* RI	304-528	M13mp18	304-528
M6	*Pvu* II-*Eco* RI	1-528	M13mp18	1-317

This table summarises the subclones used in sequencing the region of λHt363 containing the $tRNA^{Gly}$ gene. The subclones are numbered as in the sequencing strategy shown in figure 3.5.The numbers given in the insert column refer to the residues in the sequence presented in figure 3.6.

Figure 3.6. Nucleotide sequence of a 713 bp region from λHt363 containing a $tRNA^{Gly}$ gene.

The nucleotide sequence of a 713 bp region from λHt363 is shown. The position, orientation and cloverleaf secondary structure of the $tRNA^{Gly}_{GCC}$ gene are indicated. The sequence was determined for both strands with appropriate overlaps (see figure 3.5.). The limits of an identical sequence containing another human $tRNA^{Gly}_{GCC}$ gene previously published (Doran *et al* ., 1988) are indicated (*).



$tRNA^{Leu}$ (see section 3.1.) had been isolated from λHt363 as part of a 4.2 kb *Bam* HI fragment and partially sequenced. Fig. 3.7. summarises the sequencing strategy used to obtain the complete sequence of 1651 bp and table 3.3. lists the M13 subclones constructed. The sequence itself is shown in figure 3.8. The DNA was sequenced on both strands, reading through all restriction sites. Figure 3.8. also shows the position and orientation of the tRNA genes found within this region and the predicted cloverleaf secondary structures of the RNA-like strands of the genes. As mentioned in section 3.1., the sequence data previously established for this region of λHt363 appeared to be almost identical to a previously published 1650 bp sequenced region carrying the same tRNA genes on a recombinant termed λHt4 (Roy *et al* .,1982). A comparison of the complete 1651 bp sequence from λHt363 with the corresponding region on λHt4 revealed only two differences, indicated on figure 3.8. Both differences occur in the flanks of the tRNA genes rather than the coding sequences. In relation to λHt4, λHt363 has an insertion of a G residue after T_7 and a G at position 898 replacing a C at the equivalent position in λHt4. (The residue numbers refer to those given in figure 3.8.). λHt363 and λHt4 were both cloned from the same genomic library, but the restriction map of λHt4 (Buckland *et al* ., 1983) differs sufficiently from that of λHt363 to show that they represent different parts of the genome (see section 3.4. and figure 3.11.).

3.3.3. Sequencing of the $tRNA^{Arg}$ gene region.

An *in vitro* transcription assay (see section 2.5.16.2.) was performed using a pUC19 recombinant pCLU1, (so termed since it contained a tRNA gene cluster; see section 5.4.) with an insert consisting of a 4.2 kb *Bam* HI fragment from λHt363. The 1651 bp region containing the tRNA genes coding for $tRNA^{Lys}$, $tRNA^{Gln}$ and $tRNA^{Leu}$ (see figure 3.8.) is located within this 4.2 kb fragment. As described in section 5.4., transcripts were obtained from this recombinant which did not correspond to any of the three genes known to be present. Isolated restriction fragments were used in transcription assays to localise the source of the extra transcripts to a *Hin* d III- *Mst* I fragment, which had not previously been shown to hybridise to a mixed tRNA probe. This fragment was sequenced on both strands, reading through all restriction sites. The sequencing strategy used is summarised in

Figure 3.7. Strategy used for sequencing a 1651 bp region from λHt363 containing genes coding for $tRNA^{Lys}$, $tRNA^{Gln}$ and $tRNA^{Leu}$.

The restriction sites used to generate M13 subclones for sequencing a 1651 bp region from λHt363 are shown (*Aha* III (A), *Bam* HI (B), *Eco* RI (E), *Hin* d III (H), *Kpn* I (K), *Ssp* I (Ss) and *Sma* I (S)). The arrows indicate the orientation and extent of sequence determined from each subclone (see also table 3.3.). The discontinuous arrow indicates sequence determined from a subclone which was generated using a restriction site from outwith the boundaries of the figure.

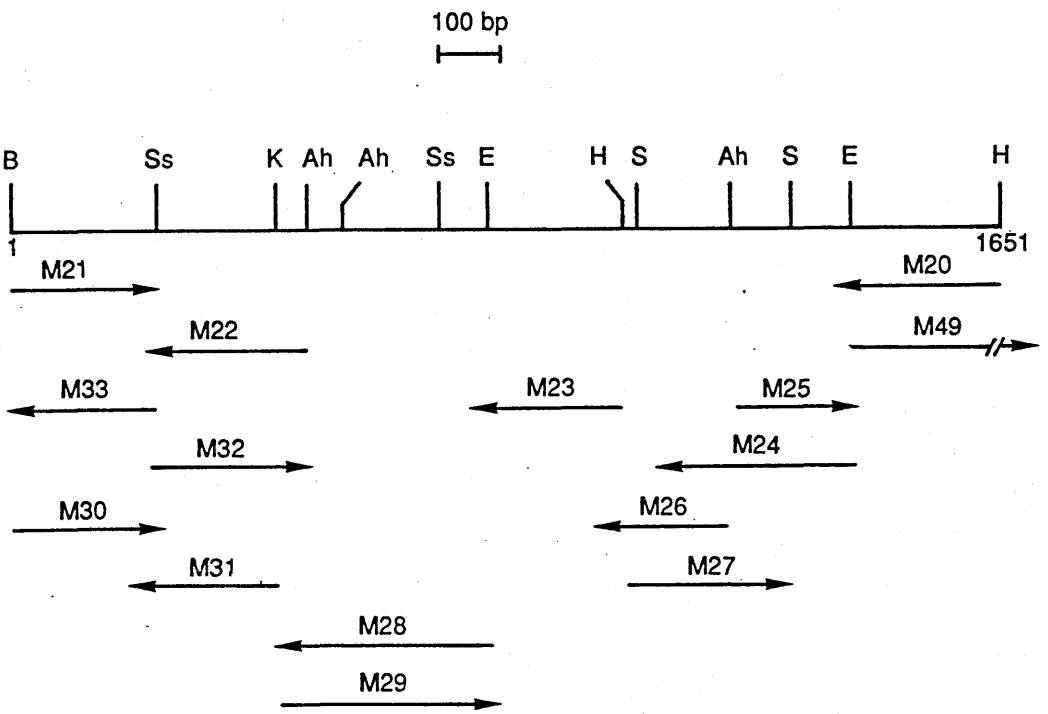


Table 3.3.

Subclones used in sequencing a 1651 bp region from λHt363 containing $tRNA^{Lys}$, $tRNA^{Gln}$ and $tRNA^{Leu}$ genes.

Subclone	Insert		Vector	Sequenced region
M20	*Hin* d III-*Hin* d III	1651-1020	M13mp18	1651-1394
M21	*Bam* HI-*Ssp* I	1-245	M13mp19	1-245
M22	*Aha* III-*Bam* HI	482-1	M13mp18	482-222
M23	*Hin* d III-*Kpn* I	1020-440	M13mp18	1020-774
M24	*Eco* RI-*Eco* RI	1406-796	M13mp18	1406-1192
M25	*Aha* III- *Eco* RI	1203-1406	M13mp18	1203-1406
M26	*Aha* III-*Eco* RI	1207-796	M13mp18	1207-984
M27	*Hin* d III-*Eco* RI	1020-1406	M13mp18	1020-1281
M28	*Eco* RI- *Bam* HI	796-1	M13mp 19	796-470
M29	*Kpn* I-*Hin* d III	440-1020	M13mp19	440-810
M30	*Bam* HI-*Kpn* I	1-440	M13mp18	1- 296
M31	*Kpn* I-*Bam* HI	440-1	M13mp19	440-203
M32	*Ssp* I-*Aha* III	245-482	M13mp18	245-482
M33	*Ssp* I-*Bam* HI	245-1	M13mp18	245-1
M49	*Eco* RI-*Eco* RI	1406-*	M13mp10	1406-*

This table summarises the subclones used in sequencing the region of λHt363 containing the tRNA genes coding for $tRNA^{Lys}$, $tRNA^{Gln}$ and $tRNA^{Leu}$. The subclones are numbered as in the sequencing strategy shown in figure 3.7. The numbers in the insert column refer to the sequence presented in figure 3.8. Subclone M49 was used to read through the terminal *Hin* d III site of this 1651 bp region into the contiguous sequence shown in figure 3.10.: * denotes a residue outwith the 1651 bp sequence shown in figure 3.8.

Figure 3.8. Nucleotide sequence of a 1651 bp region from λHt36 containing genes coding for $tRNA^{Lys}$, $tRNA^{Gln}$ and $tRNA^{Leu}$.

The nucleotide sequence of a 1651 bp region from λHt363 is shown opposite.Th positions, orientations and cloverleaf secondary structures of the tRNA genes fror within this region are indicated. The sequence was determined for both strands wit appropriate overlaps (see figure 3.7.). Residues which differ from an almos identical sequence determined earlier (Roy *et al*., 1982), are indicated (*).

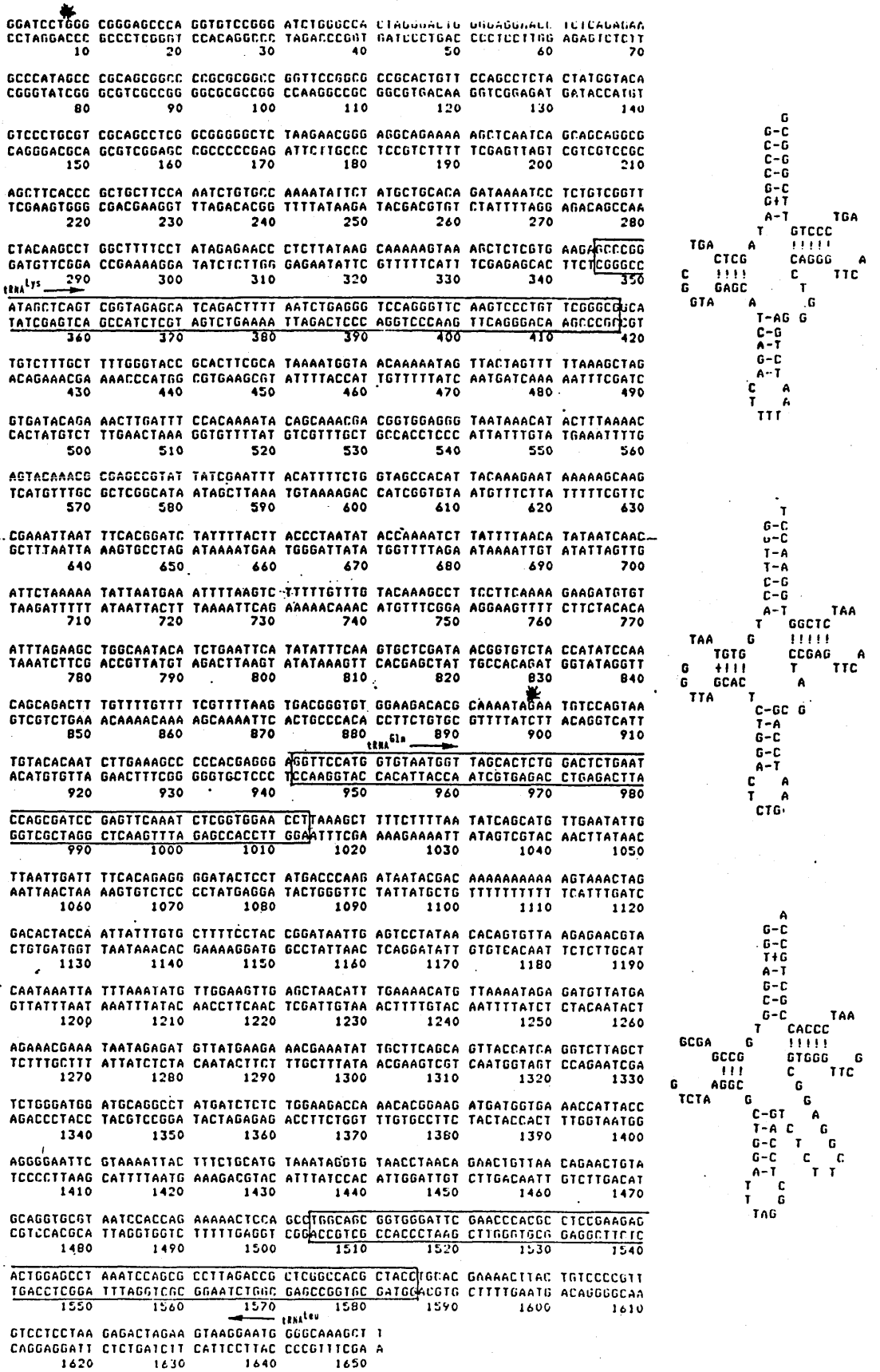


figure 3.9. and table 3.4. lists the M13 subclones constructed. The 1271 bp sequence, contiguous with the 1651 bp region described in section 3.3.2., is presented in figure 3.10. This region contains a gene coding for $tRNA^{Arg}$, the first $tRNA^{Arg}$ gene to be isolated from the human genome. The presence of this gene accounted for the extra transcripts found from the λHt363 subclone described above. The gene contains an intron, the first recorded in a $tRNA^{Arg}$ gene for any organism.

3.4. Discussion.

3.4.1. Comparison of λHt363 with other recombinants containing human tRNA genes.

λHt363 represents the largest human tRNA gene cluster cloned to date, with five tRNA genes within 7 kb of each other. The genes code for $tRNA^{Lys}_{UUU}$, $tRNA^{Gln}_{CUG}$, $tRNA^{Leu}_{UAG}$, $tRNA^{Arg}_{UCU}$ and $tRNA^{Gly}_{GCC}$. Four of the genes have exactly the same coding sequence as human tRNA genes previously reported.

The $tRNA^{Gly}_{GCC}$ gene from λHt363 has an identical coding sequence and identical 5' and 3' flanking sequences to a $tRNA^{Gly}_{GCC}$ gene previously reported and an identical coding sequence to a second $tRNA^{Gly}_{GCC}$ gene cloned by the same group (Doran *et al* ., 1988). The 5' and 3' flanks of this second $tRNA^{Gly}$ gene show no significant sequence homology to the flanking regions of the $tRNA^{Gly}$ gene from λHt363.

The 1651 bp region from λHt363 which contains the $tRNA^{Lys}$, $tRNA^{Gln}$ and $tRNA^{Leu}$ genes differs in only two residues from another human tRNA gene cluster (Roy *et al* ., 1982) on a recombinant termed λHt4. The differences are in the 5' flanking regions of the $tRNA^{Lys}$ and $tRNA^{Gln}$ genes, at distances of 336 and 44 bp respectively from the coding sequences (residues 8 and 898 in figure 3.8.). Like the recombinant λHt9 cloned by Doran *et al* ., (1988) which contains the $tRNA^{Gly}_{GCC}$ gene identical to that on λHt363, λHt4 was cloned from the same genomic library (Lawn *et al* ., 1978) as λHt363. This raises the possibility, in view of the sequence similarity between the clones, that the same genomic fragments could have been

Figure 3.9. Strategy used for sequencing a 1271 bp region of λHt363 containing a $tRNA^{Arg}$ gene.

The restriction sites used (*Hin* d III (H), *Mst* I (M), *Sau* 3A (Sa) and *Sma* I (S)) to generate M13 subclones for sequencing a 1271 bp region of λHt363 are shown. The arrows indicate direction and extent of sequence determined from each subclone (see also table 3.4.). Discontinuous arrows indicate sequence determined from subclones generated using restriction sites outwith the boundaries of the figure.

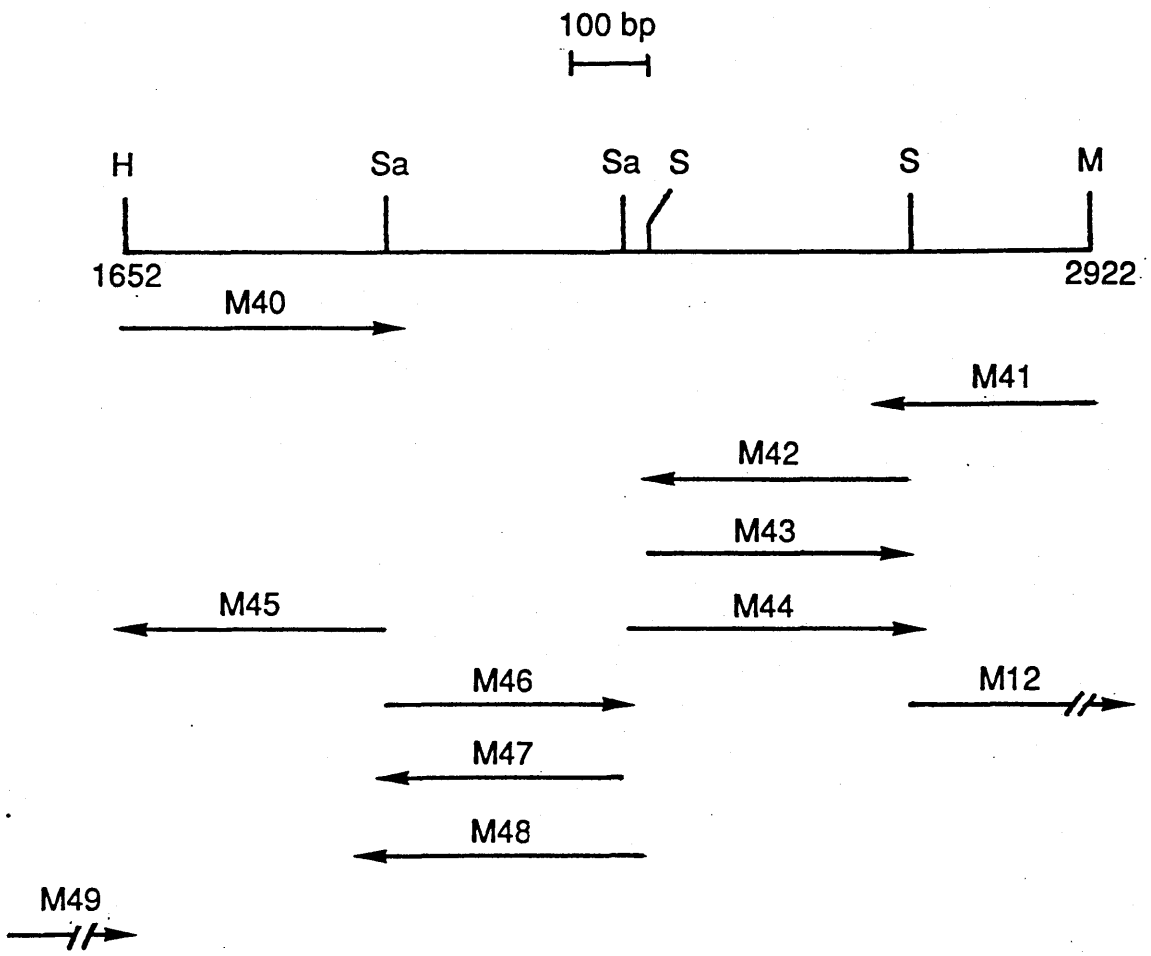


Table 3.4.

Subclones used in sequencing a 1271 bp region from λHt363 containing a gene coding for $tRNA^{Arg}$.

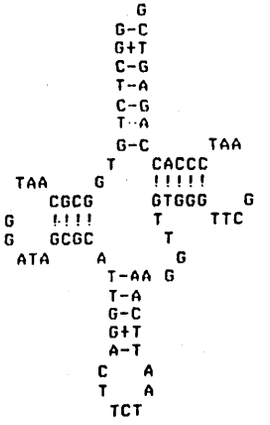
Subclone	Insert		Vector	Sequenced region
M40	*Hin* d III-*Mst* I	1652-2922	M13mp10	1652-2020
M41	*Mst* I-*Hin* d III	2922-1652	M13mp11	2922-2650
M42	*Sma* I-*Sma* I	2686-2346	M13mp11	2686-2346
M43	*Sma* I-*Sma* I	2346-2686	M13mp10	2346-2686
M44	*Sau* 3A-*Eco* RI	2302-*	M13mp18	2302-2592
M45	*Sau* 3A-*Hin* d III	1999-1652	M13mp19	1999-1652
M46	*Sau* 3A-*Sau* 3A	1999-2302	M13mp18	1999-2302
M47	*Sau* 3A-*Sau* 3A	2302-1999	M13mp18	2302-1999
M48	*Sma* I-*Hin* d III	2346-1652	M13mp11	2346-1964
M49	*Eco* RI-*Eco* RI	*-*	M13mp10	*-1662

This table summarises the subclones used for sequencing a 1271 bp region which contains the $tRNA^{Arg}$ gene from λHt363. The subclones are numbered as in the sequencing strategy shown in figure 3.9. The numbers in the insert column refer to the sequence presented in figure 3.10. Residues from outwith the sequence shown in figure 3.10. are denoted by *. The subclone M49 was used to read through the terminal *Hin* d III site of this 1271 bp region from the contiguous sequence shown in figure 3.8.

Figure 3.10. Nucleotide sequence of a 1271 bp region from λHt36 containing a $tRNA^{Arg}$ gene.

The nucleotide sequence of a 1271 bp region from λHt363 is shown opposite. Th position, orientation and cloverleaf secondary structure of the $tRNA^{Arg}$ gene a indicated. The position of the intron within this gene is also marked. The nucleotid sequence was determined for both strands with appropriate overlaps (see figu 3.9.).

```
AAGAGTGCCA AATTGATTCG TCTCCTTTTC CACATCTTTT AGACTGGAAA TGAAGAACCA TCGAGACTGT GGACCGCATA GCAGCAGATG
TTCTCACCGT TTAACTAAGC AGAGGAAAAG GTGTAGAAAA TCTGACCTTT ACTTCTTGGT AGCTCTGACA CCTGGCGTAT CGTCGTCTAC
      1661       1671       1681       1691       1701       1711       1721       1731       1741

GCAGAACAGA TAAACGAGAA GAGACAAAAA CACACAGAAA TTTACAACAA AACTAGTTTA TTTGGGTGGG ACTATATCTG GGAGTGGGGA
CGTCTTGTCT ATTTGCTCTT CTCTGTTTTT GTGTGTCTTT AAATGTTGTT TTGATCAAAT AAACCCACCC TGATATAGAC CCTCACCCCT
      1751       1761       1771       1781       1791       1801       1811       1821       1831

GCGCCTGAGC CGGAAACACC CCCCAAGCGG GCAGCCGGAG CTCTGCACTC GCTCGCTCAG CCATTTTCGG TTGCTCTCCC TCCCTGTTTT
CGCGGACTCG GCCTTTGTGG GGGGTTCGCC CGTCGGCCTC GAGACGTGAG CGAGCGAGTC GGTAAAAGCC AACGAGAGGG AGGGACAAAA
      1841       1851       1861       1871       1881       1891       1901       1911       1921

TTCGACCCAT GTGCGCCCTC TCTCTCACCT GGCTAACATG CTTTGCGCCC GCTACCTCTA CCTGGCTTGG CCTCTGAGAT CCAAACCGCG
AAGCTGGGTA CACGCGGGAG AGAGAGTGGA CCGATTGTAC GAAACGCGGG CGATGGAGAT GGACCGAACC GGAGACTCTA GGTTTGGCGC
      1931       1941       1951       1961       1971       1981       1991       2001       2011

CCCTCCTGGC ATGTCCCGCC CAATTCCCTT AGCCCCGCTC CCCAACCCTG CCCCATTCCC ACTCTAGTAC CCGTAAGCAA CAAGACGCCG
GGGAGGACCG TACAGGGCGG GTTAAGGGAA TCGGGGCGAG GGGTTGGGAC GGGGTAAGGG TGAGATCATG GGCATTCGTT GTTCTGCGGC
      2021       2031       2041       2051       2061       2071       2081       2091       2101
                      tRNA^Arg ──►                          Intron ◄──►

CCGTTCGCCG GGTGGCTCTG TGGCGCAATG GATAGCGCAT TGGACTTCTA GTGACGAATA GAGCAATTCA AAGGTTGTGG GTTCGAATCC
GGCAAGCGGC CCACCGAGAC ACCGCGTTAC CTATCGCGTA ACCTGAAGAT CACTGCTTAT CTCGTTAAGT TTCCAACACC CAAGCTTAGG
      2111       2121       2131       2141       2151       2161       2171       2181       2191

CACCAGAGTC GATTTTATTT CAATTTTTCC TTTTTATCCG TTTTTTTCTT AAAGGCCCCC ATCCCTTCTC CCCCTACCCT ATCTCTCTCT
GTGGTCTCAG CTAAAATAAA GTTAAAAAGG AAAAATAGGC AAAAAAAGAA TTTCCGGGGG TAGGGAAGAG GGGGATGGGA TAGAGAGAGA
      2201       2211       2221       2231       2241       2251       2261       2271       2281

TTCCACCTGG TTTCAATTTC GATCTCAGTT CCCGCTCTGC TCTCTTCTTT ACTTACCTGT CTGCCCCGGG GCTTGGGCCA TGGAGAGCAG
AAGGTGGACC AAAGTTAAAG CTAGAGTCAA GGGCGAGACG AGAGAAGAAA TGAATGGACA GACGGGGCCC CGAACCCGGT ACCTCTCGTC
      2291       2301       2311       2321       2331       2341       2351       2361       2371

GACTGAGGGT GGGAGACAGA AGGGAAGGGA AAGTGGGCGT GGACGTCGAG ATAAAGGCAG GCCCCTGGCT CGGGGACACA CGGGGATTTA
CTGACTCCCA CCCTCTGTCT TCCCTTCCCT TTCACCCGCA CCTGCAGCTC TATTTCCGTC CGGGGACCGA GCCCCTGTGT GCCCCTAAAT
      2381       2391       2401       2411       2421       2431       2441       2451       2461

ATAACCACTT TATTCCATGC ACTAGGGACT CGGGAAGGGA CTGGGACTGG GCTGAGGGGC AGAGAGAGTA CAATCCTAGA CGCAAGGGAG
TATTGGTGAA ATAAGGTACG TGATCCCTGA GCCCTTCCCT GACCCTGACC CGACTCCCCG TCTCTCTCAT GTTAGGATCT GCGTTCCCTC
      2471       2481       2491       2501       2511       2521       2531       2541       2551

AAGTTGGGGC AGGGAAAAGG GACAGGAACC AGGGAAATAT ATATTTATAT AGATACTCTA ATATAGACCA CATCTCACCC GCGCGCTGAG
TTCAACCCCG TCCCTTTTCC CTGTCCTTGG TCCCTTTATA TATAAATATA TCTATGAGAT TATATCTGGT GTAGAGTGGG CGCGCGACTC
      2561       2571       2581       2591       2601       2611       2621       2631       2641

CCCGCGCGCC CCACTGCCCT CCCCAACACC TGCTCGCCCG GACGCCCGGG TCCCTCTGCT GCCCTCGGGC TGGAGTCTCT ACCCCACCCC
GGGCGCGCGG GGTGACGGGA GGGGTTGTGG ACGAGCGGGC CTGCGGGCCC AGGGAGACGA CGGGAGCCCG ACCTCAGAGA TGGGGTGGGG
      2651       2661       2671       2681       2691       2701       2711       2721       2731

TAGACCCCGC CCCCACCACC CCCCAAGGCT CAGGGCCAAG GTTCTCCAGA AAGCGGGCGG CGGGGGCAGC GGGGCCTTGG GCGCCCCGTC
ATCTGGGGCG GGGGTGGTGG GGGGTTCCGA GTCCCGGTTC CAAGAGGTCT TTCGCCCGCC GCCCCCGTCG CCCCGGAACC CGCGGGGCAG
      2741       2751       2761       2771       2781       2791       2801       2811       2821

TTGTCTGTGA GGCGGCGGTG GCGGCGGCAG CAGTCCGGTG AGGGCGCCGG CGCGCCAGAA TCCCCGGCGC GCGGGGAGCA GAGGGATGGG
AACAGACACT CCGCCGCCAC CGCCGCCGTC GTCAGGCCAC TCCCGCGGCC GCGCGGTCTT AGGGGCCGCG CGCCCCTCGT CTCCCTACCC
      2831       2841       2851       2861       2871       2881       2891       2901       2911

GACCATGCGC A
CTGGTACGCG T
      2921
```



cloned more than once. The genomic library used was constructed by adding *Eco* RI linkers to genomic fragments generated by a partial digest with *Hae* III and *Alu* I, in effect generating almost random fragments which were size fractionated prior to cloning into *Eco* RI digested λCharon 4A (Maniatis *et al* .,1978). Given that overlapping fragments will be present in the library it is conceivable that the insert of λHt363 might be a genomic fragment which overlaps the inserts of λHt4 and λHt9. If this were the case then the genes shared by the different recombinants would be identical. Figure 3.11. shows a comparison of the restriction maps of λHt363, λHt4 and λHt9. This comparison shows that the restriction maps of λHt4 and λHt9 correspond to that of λHt363 in the areas close to the tRNA genes but differ sufficiently outwith these regions to confirm that the genes on λHt363 are in a different genomic environment to those on λHt4 and λHt9. (In the case of λHt363 and λHt4 there are also the sequence differences detailed above). This assumes firstly that none of the recombinants have undergone rearrangements relative to the original genomic DNA and secondly that the different copies are not allelic. These possibilities were investigated as described in sections 4.2. and 4.3.

3.4.2. The $tRNA^{Arg}$ gene from λHt363.

The $tRNA^{Arg}$ gene from λHt363 was not recorded as being present on λHt4. However, as figure 3.11. shows, this gene comes from the region of apparent homology between the two recombinants. (In fact, as will be shown in section 4.4.4.,the $tRNA^{Arg}$ gene is almost certainly present on λHt4). It is possible that the presence of this gene was not recorded because of its poor hybridisation to a tRNA probe. The $tRNA^{Arg}$ gene was only found on λHt363 through a transcription assay, having remained undetected through a series of hybridisation experiments (J.P. Goddard, A.McLaren and T.Carr, unpublished results). A mixed tRNA probe, labelled with [^{32}P]pCp (see section 2.5.11.4.) will hybridise to a restriction fragment containing the $tRNA^{Arg}$ gene from λHt363, but as shown in figure 3.12. the other tRNA genes in the cluster show a greater intensity of hybridisation. The most likely explanation for this poor hybridisation is that the presence of an intron in the gene interferes with hybridisation with a mature tRNA probe. (It has previously been reported (Hovemann *et al* ., 1980) that secondary structure formation in a

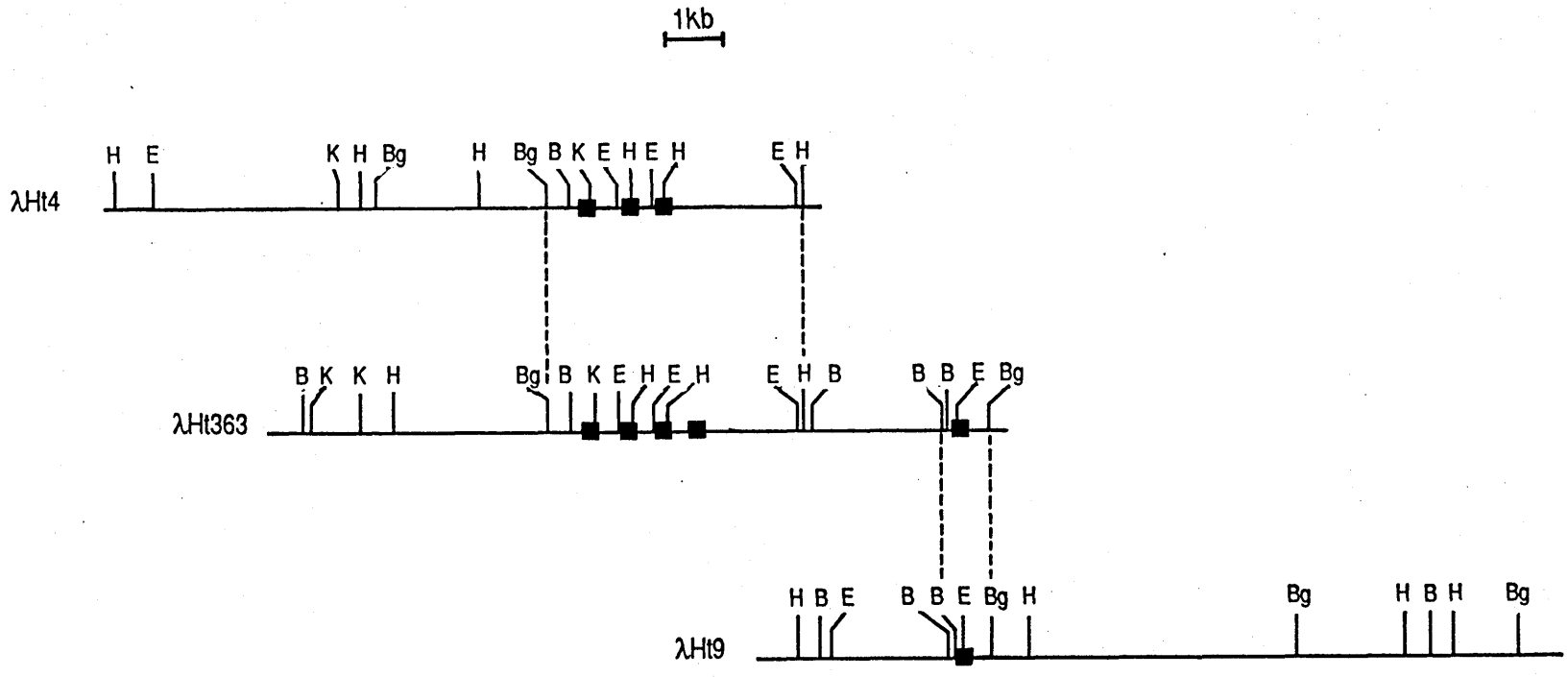


Figure 3.11. Comparison of the restriction maps of λHt363, λHt4 and λHt9.

The restriction maps of λHt363, λHt4 (Buckland *et al* ., 1983), and λHt9 (Doran *et al* ., 1988) are shown. Restriction sites are shown for *Bam* HI (B), *Bgl* II (Bg), *Eco* RI (E), *Hin* d III (H) and *Kpn* I (K). The dotted lines enclose regions with apparent similarities between the restriction maps.

tRNA gene cluster can interfere with hybridisation of a tRNA probe to one of the genes in the cluster. However,this is unlikely to be an explanation in the case of the $tRNA^{Arg}$ gene from λHt363, since the hyridisation is also poor when using the individually subcloned gene in hybridisations, as seen in figure 3.12.). This raises the possibility that the low number of human tRNA genes cloned with an intervening sequence may reflect a technical problem in their detection rather than their actual representation in the genome. (The converse may be true for tRNA gene clusters. The presence of several tRNA genes on a single recombinant may aid its selection using a mixed tRNA probe). Roy *et al* ., (1982) used a probe containing a *Xenopus* tRNA gene cluster to select the recombinant λHt4, in contrast to the mixed tRNA probe used to select λHt363 (Goddard *et al* ., 1983 and section 2.5.11.4.). The *Xenopus* tRNA gene cluster, containing eight tRNA genes, is described in section 1.4.3. It is interesting that apparently neither of these different probes hybridised well enough to the $tRNA^{Arg}$ gene to detect its presence in either copy of this tRNA gene cluster.

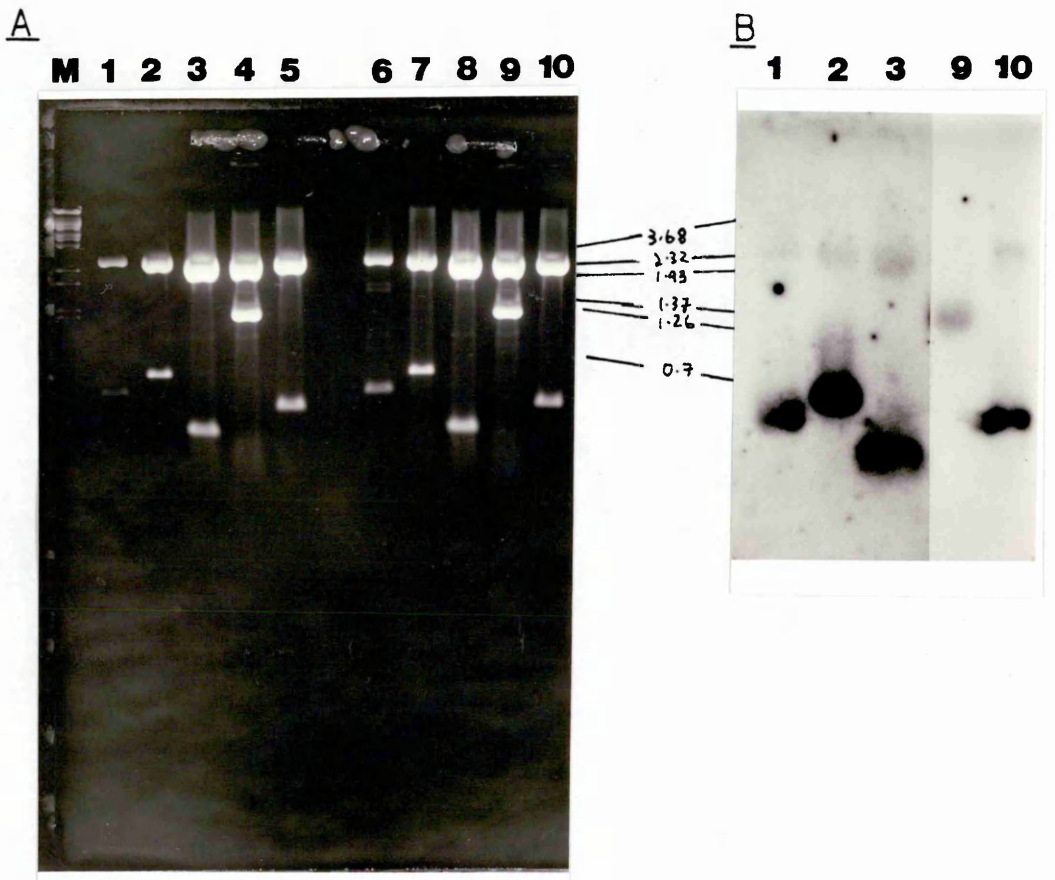
3.4.3. The G+C rich nature of λHt363.

A notable feature of the restriction map of λHt363 is the high frequency of sites for restriction enzymes which recognise G+C rich sequences. The 5.5 kb at the right hand end of the recombinant contains 10 *Sma* I sites (CCCGGG) and four *Apa* I sites (GGGCCC) as well as the unique site for the rare cutter enzyme *Not* I (GCGGCCGC) (see figure 3.3). An 841 bp *Sma* I fragment from within this region showed weak hybridisation to the mixed tRNA probe (J.P.Goddard and T.Carr, unpublished results).

This region was therefore sequenced (on one strand only) but did not contain any tRNA-like structures (data not shown). The sequence proved to be over 70% (G+C), as opposed to an average figure of 40% for mammalian DNA (Adams *et al* ., 1986). These features are consistent with the presence of a CpG-rich island, a feature characteristic of the flanking regions of mammalian genes transcribed by RNA polymerase II (Bird, 1986). This observation may be of no particular relevance for the tRNA genes themselves, since there have been no consistent reports of any bias in base composition in the genomic environment of tRNA genes,

Figure 3.12. Hybridisation of a mixed tRNA probe to pUC19 subclones containing tRNA genes from λHt363.

Figure 3.12.A Duplicate restriction digests of a set of pUC19 subclones were electrophoresed on a 1% agarose gel. Lanes 1-5 correspond to lanes 6-10 respectively. Each of the pUC19 subclones (described in table 5.1) was digested to release the insert containing the tRNA gene from the plasmid vector. Lanes 1,6 = pLYS1 digested with *Bam* HI and *Kpn* I, lanes 2,7 = pGLN1 digested with *Eco* RI, lanes 3,8 = pLEU1 digested with *Eco* RI and *Hin* d III, lanes 4,9 = pARG1 digested with *Eco* RI and *Hin* d III, lanes 5,10 = pGLY1 digested with *Bam* HI and *Hin* d III.The marker DNA fragment sizes (kb) are indicated. M = λ DNA/*Bst* E II. Figure 3.12.B shows two parts of the autoradiograph obtained from hybridisation of the blotted gel to a mixed tRNA probe. The lanes are numbered as in figure 3.12.A. The marker DNA fragment sizes (kb) are indicated.



although Gamulin *et al* ., (1983) reported the characterisation of yeast tRNA genes in an A+T rich environment. It is however possible that the highly G+C rich nature of the 841 bp *Sma* I fragment (which overlaps with residues 2686-2922 of the sequence shown in figure 3.10.) leads to its spurious hybridisation to a mixed tRNA probe. This could arise since the coding regions of a number of tRNA genes are relatively G+C rich (Sprinzl *et al* ., 1989).

CHAPTER FOUR

ASPECTS OF THE GENOMIC ORGANISATION OF THE tRNA GENE CLUSTER FROM λHt363.

4.1. Background and aims

The results discussed in section 3.4. suggest that the cluster of tRNA genes found on λHt363 occurs in at least two copies in the human genome, despite the fact that *in situ* hybridisation experiments (see section 4.4.3.) have shown that probes from λHt363 only hybridise to one chromosomal location. This interpretation depends on the integrity of the inserts of the relevant recombinant phages λHt363, λHt4 and λHt9. If any of these inserts has undergone rearrangements relative to the original genomic DNA then a comparison of restriction maps would be meaningless. Evidence that the insert of λHt363 has not undergone any rearrangments was provided by two lines of investigation. The first involved Southern hybridisations of labelled restriction fragments containing tRNA genes to genomic DNA digested with a range of restriction enzymes. The second method made use of the polymerase chain reaction (PCR) using primers specific for different parts of the tRNA gene cluster with genomic DNA as the target for amplification. These experiments, including the cloning of two human $tRNA^{Gln}$ genes as a by-product of the main study, are described in sections 4.2. and 4.3.. Southern hybridisations were also used in order to attempt to determine the copy number of the cluster as a whole and of the individual genes if they occur outwith the cluster.

The versatile PCR is useful for a number of techniques in which genomic sequences can be analysed without their prior cloning (see section 2.5.15.). Section 4.3.2. describes the use of the PCR in an attempt to determine the nature of the intervening sequence in genomic copies of the $tRNA^{Arg}$ gene family.

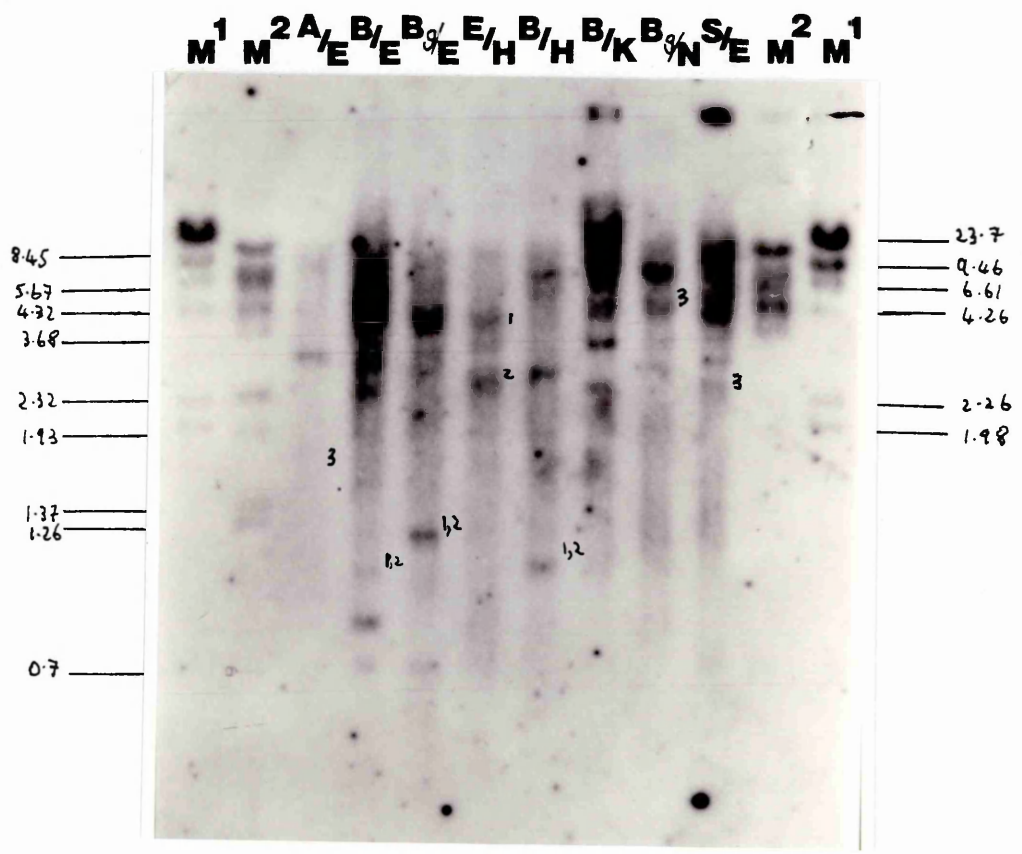
4.2. Genomic Southern hybridisations.

4.2.1. The use of restriction fragments from λHt363 as probes.

Restriction fragments from λHt363 were labelled with α-[^{32}P]-dCTP as described in section 2.5.11.1. The labelled restriction fragments were hybridised to genomic DNA samples digested with different restriction endonucleases, separated by agarose gel electrophoresis and fixed on nylon filters (see sections 2.5.6., 2.5.7. and 2.5.10.). Figure 4.1. shows the results of a hybridisation experiment using a 440 bp *Bam* HI/*Kpn* I fragment from λHt363 containing the tRNA^{Lys} gene (corresponding to sequence residues 1 to 440 in figure 3.8.). A number of fragments in each lane have hybridised to the probe. The bands marked (1,2) (*Bam* HI/*Eco* RI 0.85 kb, *Bgl* II/*Eco* RI 1.3 kb, *Hin* d III/*Bam* HI 1 kb) are those corresponding in size to restriction fragments predicted on the basis of the restriction maps of λHt363 and λHt4, both of which contain the 440 bp fragment used as a probe. The 2.5 kb *Eco* RI/*Hin* d III band marked (2) corresponds to a restriction fragment size predicted from the restriction map of λHt4 but not from that of λHt363. The band corresponding to the *Eco* RI/*Hin* d III 4 kb fragment predicted on the basis of the map of λHt363 is marked (1). The *Eco* RI/*Hin* d III genomic digest is the only one shown for which different hybridising bands can be predicted for λHt363 and λHt4. For the other digests shown, either the recombinants share the relevant restriction sites and therefore the same predicted fragment sizes, or restriction enzymes have been used which are not shown on the published map of λHt4 (Buckland *et al* ., 1983). Bands marked (3) (*Apa* I/*Eco* RI 1.5 kb, *Not* I/*Bgl* II 4.8 kb and *Sma* I/*Eco* RI 2.3 kb) are those which correspond to sites predicted by the restriction map of λHt363 for which no prediction was possible from the map of λHt4. In similar experiments, restriction fragment probes from elsewhere in the insert of λHt363 also detected bands from genomic digests which corresponded in size to fragments consistent with the restriction maps of the inserts of λHt363, λHt4 (Buckland *et al* ., 1983) and λHt9 (Doran *et al* .,1988).

Figure 5.2. Hybridisation of a restriction fragment containing the $tRNA^{Lys}$ gene from λHt363 to restriction digests of human genomic DNA.

This figure shows the hybridisation of a labelled 440 bp *Bam* H I-*Kpn* I fragment (corresponding to the insert of pLYS1: see table 5.1.) to a series of double restriction digests of human genomic DNA. The restriction enzymes used were *Apa* I (A), *Bam* HI (B), *Bgl* II (Bg), *Eco* RI (E), *Hin* d III (H), *Kpn* I (K), *Not* I (N) and *Sma* I (S). Restriction fragments of genomic DNA were electrophoresed on a 0.8 % agarose gel before being blotted onto a nylon membrane. Genomic DNA fragments of the sizes predicted from the restriction maps of λHt363 and λHt4 (Buckland *et al* ., 1983) are labelled 1, 2 or 3 to the right of the bands in question. Restriction fragments corresponding to the map of λHt363 are labelled 1, and those fragments corresponding to the map of λHt4 are labelled 2. Limited restriction data for λHt4 precluded assignment of some bands; bands labelled 3 are those which correspond to fragment sizes predicted on the basis of the restriction map of λHt363 for which no prediction was possible from the map of λHt4. No fragments are indicated in the *Bam* HI/*Kpn* I lane: fragments of the expected hybridising fragment size, 440 bp, would have run off the gel. Marker DNA fragments were end-labelled with [^{32}P]dATP. The marker fragment sizes (kb) are indicated. M1 = λ DNA/*Hin* d III, M2 = λ DNA/*Bst* E II.



4.2.2. The use of PCR generated probes.

The coding sequence of tRNA genes spans only 70-95 bp. It follows that for restriction fragments of greater than 180 bp carrying a tRNA gene the majority of sequence represented will be outside the coding region of the gene. The consequence of this is that when such restriction fragments are used as probes against digests of genomic DNA as described in section 4.2.1. the pattern of hybridisation will reflect the nature of the sequence flanking the tRNA gene rather than the gene itself. In order to make a rough estimate of the genomic copy number of the tRNA genes from λHt363, it was desirable to construct a probe for each gene consisting only of the coding region. Such a probe in theory should recognise all of the genomic copies of the gene in a Southern hybridisation. A rough estimate of minimum copy number can be made by counting the number of bands detected in genomic DNAs cut with different restriction enzymes.

The most convenient method found of constructing such gene-specific probes for the five tRNA genes in the cluster was to use the PCR. Oligonucleotide primers were designed for the 5' and 3' ends of each gene (see table 4.1. and section 4.3.) and the PCR used to amplify the gene sequence as described in section 2.5.15.1. The amplified fragments (shown in figure 4.2.) were labelled as described in section 2.5.11.1. and hybridised to genomic DNA . Figure 4.3. shows the results of a hybridisation experiment using a labelled probe consisting of the $tRNA^{Arg}$ gene sequence. The marked bands (*Bam* HI 4.2 kb, *Bgl* II 8.1 kb, *Eco* RI 2.6 kb and *Hin* d III 2.5 kb) show restriction fragments corresponding in size to fragments which would be predicted to hybridise to this probe on the basis of the restriction map of λHt363.

Similar hybridisation experiments were performed for each of the tRNA genes from λHt363. The largest number of bands that could be detected in a single genomic digest was taken as an estimate of minimum copy number. On the basis of these results, all of the genes from λHt363 appear to be present in multiple copies in the genome, and in different genomic environments. The minimum copy numbers estimated were $tRNA^{Lys}$ = 9, $tRNA^{Gln}$ = 10, $tRNA^{Leu}$ = 10, $tRNA^{Arg}$ = 7 and $tRNA^{Gly}$ = 8.

Table 4.1.

Oligonucleotide primers used in polymerase chain reactions to amplify gene specific probes.

Primer	Position in tRNA gene	Sequence reference
A	5' end of $tRNA^{Lys}$ gene (residues 1-17)	345-361 figure 3.8.
B	3' end of $tRNA^{Lys}$ gene (residues 73-57)	417-401 figure 3.8.
C	5' end of $tRNA^{Gln}$ gene (residues 1-17)	942-958 figure 3.8.
D	3' end of $tRNA^{Gln}$ gene (residues 72-56)	1013-997 figure 3.8.
E	5' end of $tRNA^{Leu}$ gene (residues 1-17)	1585-1569 figure 3.8.
F	3' end of $tRNA^{Leu}$ gene (residues 82-66)	1504-1520 figure 3.8.
G	5' end of $tRNA^{Arg}$ gene (residues 1-17)	2115-2131 figure 3.10.
H	3' end of $tRNA^{Arg}$ gene (residues 88-72)	2202-2186 figure 3.10.
I	5' end of $tRNA^{Gly}$ gene (residues 1-17)	503-519 figure 3.6.
J	3' end of $tRNA^{Gly}$ gene (residues 71-55)	573-557 figure 3.6.

This table summarises the oligonucleotide primers constructed for use in the polymerase chain reaction and lists the positions of the 17-mer oligonucleotides within the tRNA genes. Primers for the 5' ends of the tRNA genes have the same sequence as the tRNA-like strand so it follows that for the polymerase chain reaction primers for the 3' ends of the tRNA genes must have the same sequence as the coding strand. The sequences of these primers can be found in the sequence data presented in the figures listed above.

Figure 4.2. Amplification of tRNA genes from λHt363.

Oligonucleotide primers specific for amplification of each of the five tRNA genes from λHt363 were used with target λHt363 DNA in the polymerase chain reaction. The primers used are listed in table 4.1. The amplified products are shown after electrophoresis through a 10% polyacrylamide gel and ethidium bromide staining. The numbered lanes are 1 = $tRNA^{Lys}$ gene, 2 = $tRNA^{Gln}$ gene, 3 = $tRNA^{Leu}$ gene, 4 = $tRNA^{Arg}$ gene, 5 = $tRNA^{Gly}$ gene. The residual primers are visible at the bottom of the gel. Marker DNA fragment sizes (bp) are indicated. M = M13mp18 DNA/*Hin* f I.

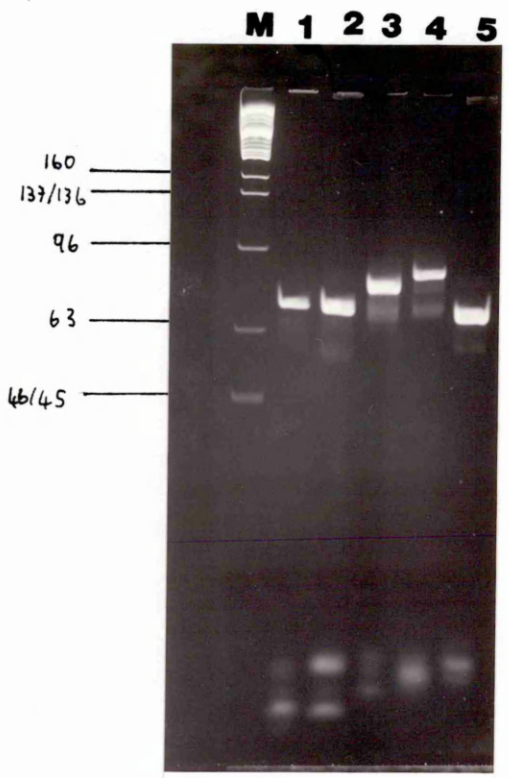
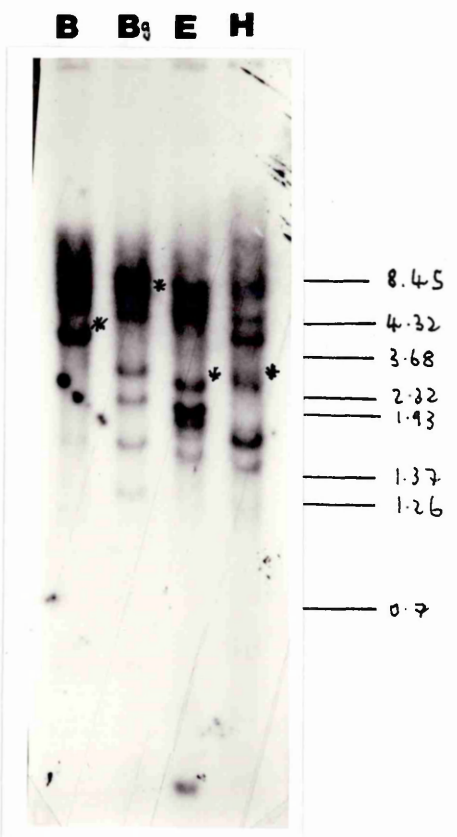


Figure 4.3. Hybridisation of the λHt363 tRNA^{Arg} gene to restriction digests of human genomic DNA.

This autoradiograph shows the result of a hybridisation of the tRNA^{Arg} gene amplified from λHt363 by the PCR to a Southern blot of restriction digests of human genomic DNA. The restriction digests used were *Bam* HI (B), *Bgl* II (Bg), *Eco* RI (E) and *Hin* d III (H). Genomic restriction fragments of the sizes predicted by the restriction map of λHt363 are indicated (*) to the right of the bands. Marker fragment sizes (λ DNA/ *Bst* E II) are shown (in kb).



4.2.3. Estimation of the tRNA gene cluster copy number.

An estimate of the copy number for the tRNA gene cluster from λHt363 was made by comparing the intensity of hybridisation of a restriction fragment containing part of the cluster to genomic DNA with the intensity of hybridisation to a series of standards. Three different amounts of genomic DNA (4, 2 and 1 μg) were digested with *Bam* HI. DNA from the recombinant λHt363, also digested with *Bam* HI, was used as standards in amounts ranging from 10-100 pg. The digests were run in parallel on a 0.8% agarose gel with suitable size markers (see figure 4.4.). The digests of λHt363 DNA were electrophoresed in the presence of undigested λ carrier DNA. The separated DNA fragments were blotted onto a nylon membrane as described in section 2.5.10.1. and hybridised to an α-[^{32}P]-dCTP labelled probe. The probe used was a 4.2 kb *Bam* HI fragment from λHt363 containing the genes coding for $tRNA^{Lys}$, $tRNA^{Gln}$, $tRNA^{Leu}$ and $tRNA^{Arg}$ (see figures 3.3., 3.8. and 3.10.). Under the hybridisation conditions used, the probe hybridised only to a 4.2 kb band in each of digests (see figure 4.4.). The intensity of the 4.2 kb bands on the autoradiograph was quantified as described in section 2.5.13. An example of a densitometer tracing is shown in figure 4.5. together with the graph produced from the densitometer readings.The results were processed as follows. Using the standard curve shown in figure 4.5., the intensity of hybridisation for a given amount of genomic DNA (A) was related to the amount of λHt363 DNA giving a corresponding densitometer reading (B). The gene cluster copy number could then be estimated using this equation:

$$\frac{A \times \text{cluster copy number}}{\text{m.w. human DNA}} = \frac{B}{\text{m.w. } \lambda Ht363}$$

The results obtained with three different amounts of genomic DNA (1, 2 and 4 μg) gave estimated copy numbers for the tRNA gene cluster as 2.6, 3.1 and 2.5 respectively, giving an average copy number of 2.7 per haploid genome.

A similar experiment was undertaken using a *Bam* HI-*Stu* I restriction fragment of 356 bp as a probe. This restriction fragment corresponds to the insert of pGLY2 (see figure 5.9.), containing the $tRNA^{Gly}$ gene from λHt363 together with 285 bp of flanking sequence. This sequence is identical with the corresponding region from λHt9 for at least as far as the published sequence for this recombinant (Doran

Figure 4.4. Hybridisation of the 4.2 kb *Bam* HI fragment from λHt363 to *Bam* HI digested human genomic and λHt363 DNA.

Figure 4.4. A shows three different amounts of human genomic DNA digested with *Bam* HI and electrophoresed in parallel with five different amounts of λHt363 DNA also digested with *Bam* HI. Lanes 1-3 genomic DNA (1 = 4 μg, 2 = 2 μg, 3 = 1 μg) Lanes 4-8 λHt363 DNA (4 = 10 pg, 5 = 25 pg, 6 = 50 pg, 7 = 75 pg, 8 = 100 pg). The λHt363 digests were electrophoresed in the presence of λ DNA carrier. Marker DNA fragment sizes (kb) are indicated. M = λ DNA/*Eco* RI+*Hin* d III.
B shows the autoradiograph derived from hybridisation of a 4.2 kb *Bam* HI probe containing the $tRNA^{Lys}$, $tRNA^{Gln}$, $tRNA^{Leu}$ and $tRNA^{Arg}$ genes from λHt363 to a blot of the gel described in A. Marker DNA fragment sizes are indicated, as are the positions of the wells (W).

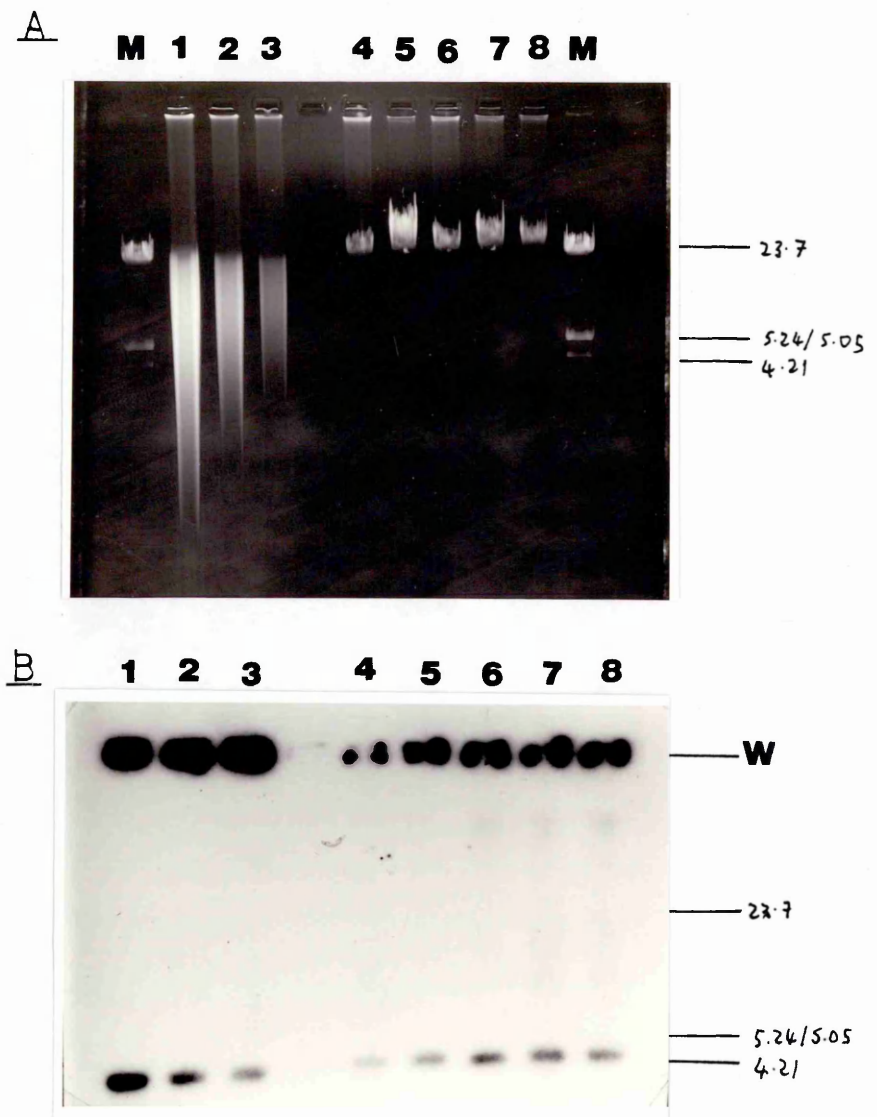
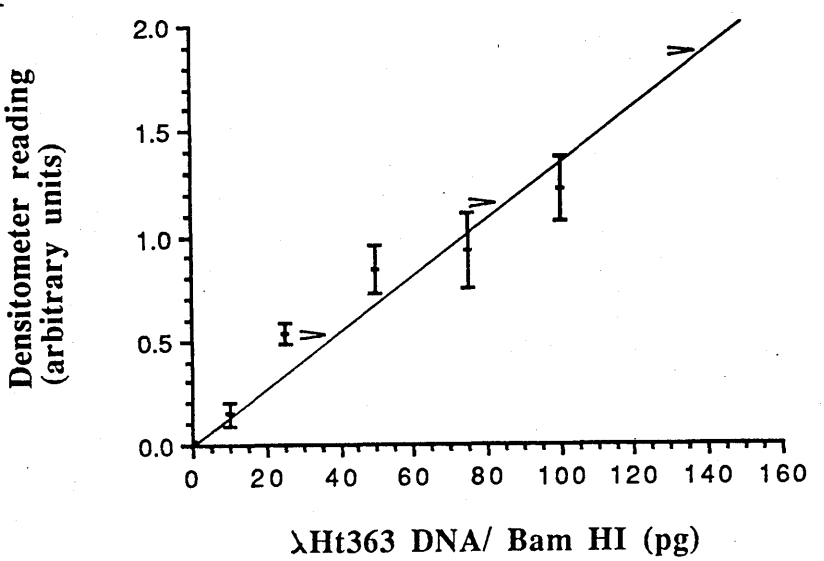


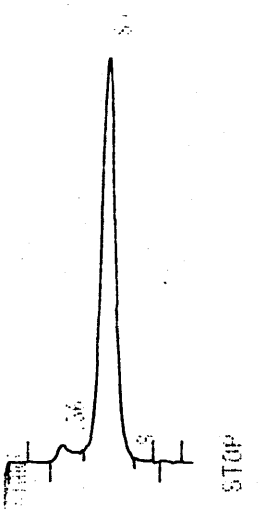
Figure 4.5. Estimation of the copy number of the tRNA gene cluster.

A standard curve derived from λHt363/*Bam* HI digestions is shown in A. Error bars indicate ± one standard deviation for four densitometric readings from the autoradiograph shown in figure 4.4. B. Arrows represent the densitometric readings obtained from the 4.2 kb bands from 1, 2 and 4 μg of genomic DNA digested with *Bam* HI. The equation for the estimation of copy number is given in section 4.2.3. Figure 4.5.B shows an example of a densitometric trace.

A



B



et al ., 1988) overlaps with that from λHt363 (343 bp; see figure 3.6.). The probe was labelled and hybridised to a filter similar to that described above, except that the genomic and λHt363 DNAs were digested with *Bam* HI and*Stu* I. However, this probe appeared to detect several bands in genomic digests, and no accurate quantification of hybridisation was possible.

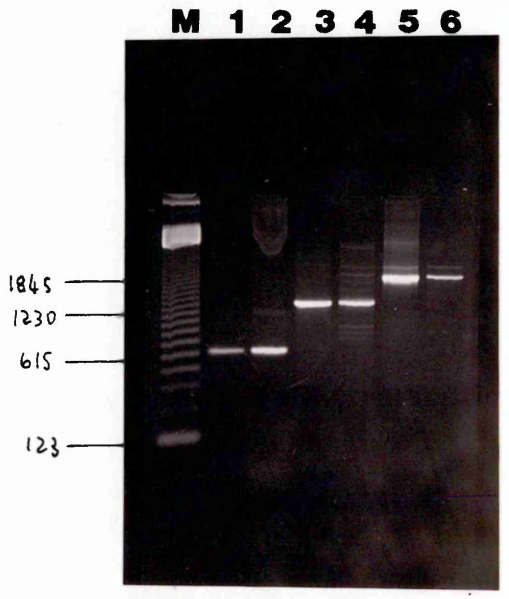
4.3. Use of the polymerase chain reaction (PCR) in the characterisation of the human tRNA gene cluster.

4.3.1. Comparison of fragments amplified from λHt363 and from genomic DNA.

Oligonucleotides for use as specific primers for different regions of λHt363 were synthesised as described in section 2.4.4. Table 4.1. lists the primers used with reference to the sequence data presented in Chapter three. These primers were used in different combinations using either λHt363 (10 ng) or genomic DNA (0.2 μg) as a target to ascertain whether the same sizes of fragments were amplified. This served as another method to look for rearrangements within the regions coding for the tRNA genes (see sections 4.2. and 4.2.1.). Among the combinations of primers used were the three which yielded the amplified fragments shown in figure 4.6. using either λHt363 or genomic DNA as a target. The first primer combination, C + F, gives an amplified fragment with the PCR of 644 bp corresponding to the sequence between the $tRNA^{Gln}$ gene 5' and the $tRNA^{Leu}$ gene 5'. The second combination, C + H, gives an amplified fragment of 1260 bp corresponding to the sequence between the $tRNA^{Gln}$ gene 5' and the $tRNA^{Arg}$ gene 3'. Finally, the primer combination A + H gives an amplified fragment of 1857 bp corresponding to the sequence between the $tRNA^{Lys}$ gene 5' and the $tRNA^{Arg}$ gene 3'. As shown in figure 4.6., λHt363 DNA and genomic DNA targets both gave rise to amplified fragments of the same size. The $tRNA^{Gly}$ gene proved to be too far from the $tRNA^{Arg}$ gene (about 5 kb) to amplify the region between the genes using the PCR with either λHt363 or genomic DNA as a target. This may be due to the G+C rich nature of the sequence between these two genes (see section 3.4.3.). The conditions for the PCR were modified to attempt to overcome this problem by replacing dGTP in the reaction mix with 7-deaza-dGTP

Figure 4.6. Comparison of fragments amplified by the polymerase chain reaction using λHt363 or genomic DNA as targets for the same primers.

A set of polymerase chain reaction products were analysed by electrophoresis through a 1% agarose gel. The oligonucleotide primer combinations used for PCR amplification (see table 4.1.) were as follows: lanes 1 and 2, primer C + primer F, lanes 3 and 4, primer C + primer H, lanes 5 and 6, primer A and primer H. Target DNA for the PCR was either λHt363 (lanes 1, 3 and 5) or human genomic DNA (lanes 2, 4 and 6). Marker DNA fragment sizes are indicated (bp). M = 123 bp ladder (BRL Ltd.).



(see section 2.5.15.1.) but no discrete products could be observed even when using λHt363 DNA as the target for amplification.

In all of the cases when amplified fragments were generated by the PCR, the fragments were of the same size whether genomic DNA or λHt363 DNA were the targets. This supports the results obtained from Southern hybridisations (sections 4.2.1. and 4.2.2.) suggesting that the insert of λHt363 has not undergone any rearrangement relative to the original genomic DNA.

4.3.2. Amplification of genomic copies of the $tRNA^{Arg}$ gene family.

Oligonucleotide primers G and H, specific for the 5' and 3' ends of the $tRNA^{Arg}$ gene, were constructed as described (see section 2.4.4. and table 4.1.). These primers were then used to amplify target genomic DNA. The rationale behind this experiment was that if genomic copies of the $tRNA^{Arg}$ gene exist without an intron then two different sizes of fragment should be amplified. (Shortly after the completion of this work, Green *et al* ., (1990) described the cloning of intron-containing $tRNA^{Leu}$ genes using a PCR technique in much the same way). The fragments would be sufficiently different in size (73 and 88 bp) to be resolved on a 10% acrylamide gel. In the event, only one fragment of 88 bp was amplified from genomic DNA. This would suggest that all of the genomic copies of $tRNA^{Arg}$ genes recognised by the primers specific for the $tRNA^{Arg}$ gene from λHt363 have an intron. The band amplified from genomic DNA is indistinguishable in size and sharpness from that amplified using λHt363 as the target DNA (see figure 4.2.). This suggests that within the resolving limit of the gels used (approximately ± 1 bp) the introns in the genomic copies of all of the $tRNA^{Arg}$ genes appear to be of the same size.

Further information on the nature of the introns in genomic copies of the $tRNA^{Arg}$ gene family was obtained using a direct sequencing method involving asymmetric PCR (see section 2.5.15.2.). The substrate for the sequencing reaction was predominantly single stranded DNA amplified from genomic DNA using the oligonucleotide primers G and H specific for the $tRNA^{Arg}$ gene. This DNA should consist of a mixed population of amplified fragments corresponding to all of the

genomic members of the $tRNA^{Arg}$ gene family. The asymmetric PCR technique can be used to give sequence information from both strands of the amplified DNA. Figure 4.7. shows the results of a sequencing reaction using genomic DNA, in which the $tRNA^{Arg}$ gene coding strand was preferentially amplified, as template and oligonucleotide H as primer. The sequence can be read unambiguously as far as the end of the coding sequence of the gene. If differences occur in either the length or the sequence of the intron in any of the copies of the $tRNA^{Arg}$ gene then the sequence from a mixed population of amplified fragments would show these differences. The results obtained by sequencing both strands of the $tRNA^{Arg}$ genes amplified from genomic DNA suggest that all of the members of this gene family have identical introns since a unique sequence can be read for the whole length of the coding region. Bands can be seen in the autoradiograph shown in figure 4.7. extending further than the sequence corresponding to the coding region of the $tRNA^{Arg}$ gene. These artifacts presumably arise through priming of DNA synthesis in the sequencing reaction from fragments amplified incorrectly in the asymmetric PCR, since fragments amplified specifically should only extend as far as the end of the primers used.

4.3.3. Amplification and cloning of two human $tRNA^{Gln}$ genes.

The PCR was used to generate gene-specific probes for each of the tRNA genes from λHt363 as described in section 4.2.2. The oligonucleotide primers C and D (see table 4.1.) specific for the 5' and 3' ends of the $tRNA^{Gln}$ gene, were used in one set of PCR experiments. When λHt363 or the plasmid construct pGLN1 (see table 5.1.) containing the $tRNA^{Gln}$ gene from λHt363 were used as the target DNA, only one band was amplified. When the same primers were used with genomic DNA as the target, two amplified bands resulted (see figure 4.8.). The smaller band amplified in all of the reactions corresponds in size to the $tRNA^{Gln}$ gene coding sequence (72 bp). The larger band found when genomic DNA was the target for amplification was estimated to be 620 bp ± 20 bp. This 620 bp band was eluted from a low melting point agarose gel as described in section 2.5.7.2., and itself used as a target for a PCR using the two primers described above. This fragment again generated two amplified bands, of 72 bp and 620 bp.

These results could be explained if there existed in the genome two $tRNA^{Gln}$ genes within 620 bp of each other. A proportion of the amplified PCR products

Figure 4.7. Direct sequencing of tRNA^{Arg} gene copies amplified from human genomic DNA.

The oligonucleotide primers G and H (table 4.1) were used in asymmetric polymerase chain reactions to amplify predominantly single stranded DNA corresponding to the coding sequences of a mixed population of genomic tRNA^{Arg} genes. The autoradiograph shows the products of one such reaction, sequenced by the Sanger chain termination method using oligonucleotide H as primer and analysed by polyacrylamide gel electrophoresis in parallel with products from a standard sequence analysis of M13mp18 (M). The region corresponding to the tRNA^{Arg} gene, including the intron, is indicated.

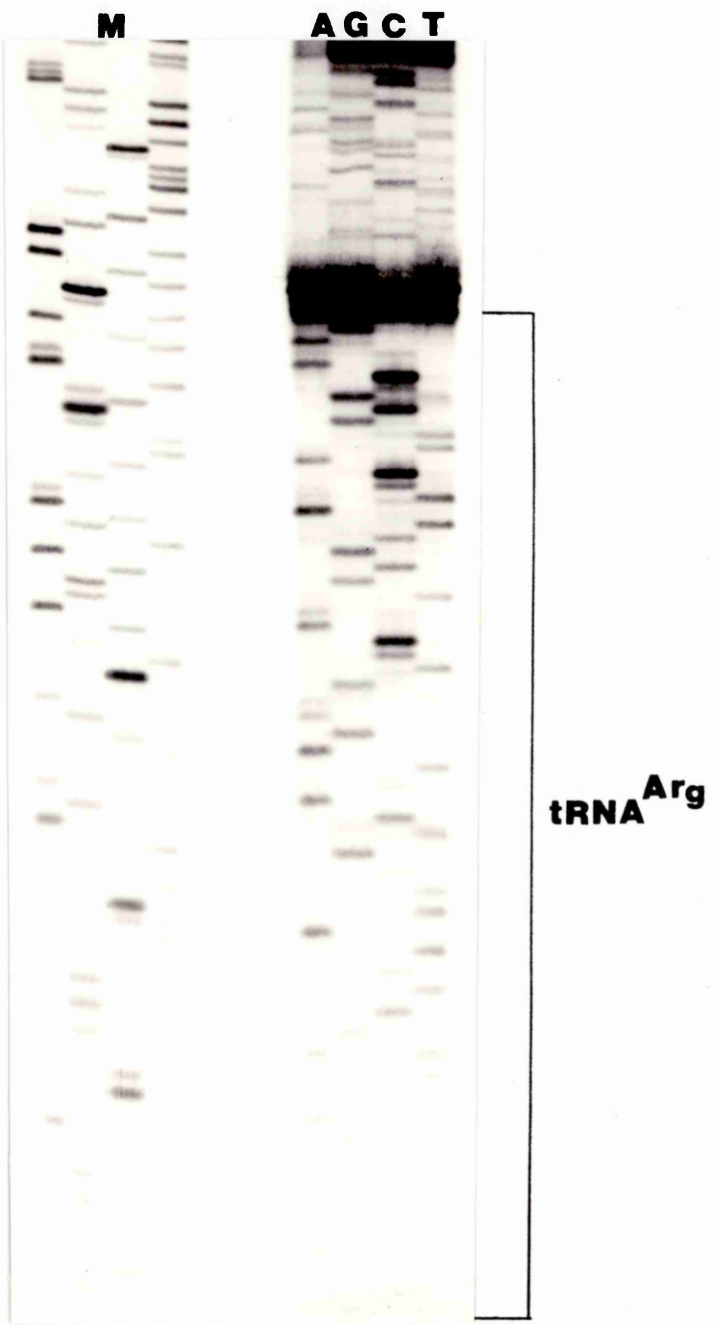
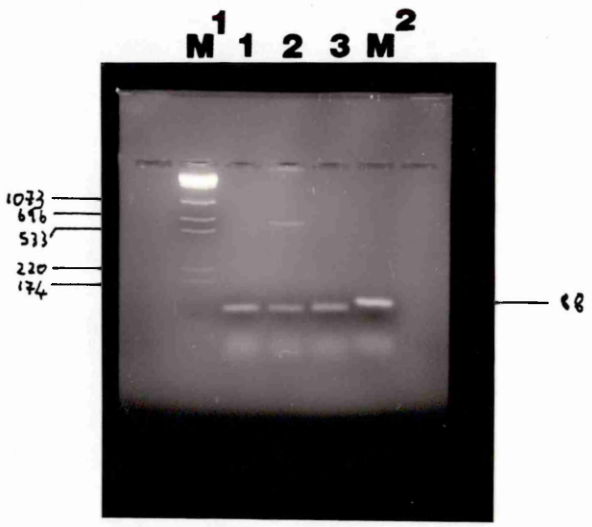


Figure 4.8. Polymerase chain reaction amplification of two genomic DNA fragments using primers specific for a $tRNA^{Gln}$ gene.

The oligonucleotide primers C and D were used in polymerase chain reactions for three different target DNAs and the products analysed by electrophoresis through a 2% agarose gel. The target DNAs were : lane 1, pGLN1 (see table 5.1); lane 2, human genomic DNA; lane 3, λHt363. Marker DNA fragment sizes are indicated (bp). M1 = λ DNA/*Dra* I, M2 = PCR amplified $tRNA^{Arg}$ gene DNA.



would be expected to correspond in length to the distance from the 5' end of one gene to the 3' end of the other. Since neither primer alone gave any amplified product with any of the target DNAs this explanation is possible only if these two genes were in the same orientation.

Sufficient amounts of the 620 bp amplified band were eluted from low melting point agarose gels to allow digestion of the fragment with restriction enzymes and to produce a preliminary restriction map. The restriction fragments listed in table 4.2. were cloned into M13 vectors and sequenced as described in section 2.5.14. The sequencing strategy used is summarised in figure 4.9. and the nucleotide sequence shown in figure 4.10. The sequence was determined on both strands, reading through all restriction sites. Two identical human $tRNA^{Gln}$ genes, in the same orientation, were found on this 623 bp sequence. The genes have completely different flanking DNA to that of the $tRNA^{Gln}$ gene from λHt363 (although only one flank of each gene is present on this 623 bp fragment). Both genes differ in a single residue from the coding sequence of the $tRNA^{Gln}$ gene from λHt363. The difference, indicated on figure 4.10., is in the anticodon. These novel human tRNA genes code for $tRNA^{Gln}_{UUG}$ rather than $tRNA^{Gln}_{CUG}$. It should be noted that since this fragment was isolated using the PCR, the residues corresponding to the positions of the primers (1-17 and 607-623 on figure 4.10.) will not necessarily have the sequence present in the original genomic DNA.

4.4. Discussion.

4.4.1. The integrity of the insert of λHt363.

Human tRNA genes are members of complex multigene families, and as yet no clear pattern of genomic organisation has been established for tRNA genes in humans or in any other mammalian system. The results discussed in section 3.4. suggest that the genes from the tRNA gene cluster represented on λHt363 occur in at least two copies in the genome. This interpretation depends firstly on the integrity of the inserts of λHt363, λHt4 and λHt9. If any rearrangements have occured during the cloning of these recombinants then comparisons of their restriction maps will be meaningless. Secondly, this interpretation of the results depends on the

Table 4.2.

Subclones used in sequencing a 623 bp fragment of human genomic DNA containing two tRNA^{Gln} genes.

Subclone	Insert		Vector	Sequenced region
M50	*Sau* 3A-*Sau* 3A	136-45	M13mp18	136-45
M51	*Sau* 3A-*Sau* 3A	45-136	M13mp18	45-136
M52	*Pst* I-*Sau* 3A	187-136	M13mp10	187-136
M53	*Pst* I-*Sau* 3A	187-596	M13mp10	187-540
M54	*Sau* 3A-*Pst* I	596-187	M13mp11	596-203
M55	*Sau* 3A-*Pst* I	136-187	M13mp11	136-187
M56	*Pst* I-L.end	187-1	M13mp10	187-1
M57	*Sau* 3A-*Sau* 3A	596-136	M13mp18	596-176
M58	L.end-*Pst* I	1-187	M13mp11	1-187
M59	*Pst* I-R.end	187-623	M13mp10	187-602
M60	R.end-*Sau* 3A	623-596	M13mp11	623-596

This table summarises the subclones used to sequence a 623 bp fragment of human genomic DNA containing two tRNA^{Gln} genes. The subclones are numbered as in the sequencing strategy shown in figure 4.9. The numbers in the insert column refer to the sequence presented in figure 4.10. L.end and R.end denote the blunt ends (at the left and right of the fragment as shown in figure 4.9.) which occured as a result of the method used in cloning this fragment (see section 4.3.3.).

Figure 4.9. Strategy used for sequencing a PCR amplified 623 bp fragment of human genomic DNA containing two $tRNA^{Gln}$ genes.

This figure summarises the restriction sites used (*Pst* I (Ps) and *Sau* 3A (Sa)) to generate M13 subclones for sequencing a PCR amplified 623 bp fragment of human genomic DNA. BE denotes blunt end. The arrows indicate the direction and extent of sequence determined from each subclone (see also table 4.2.).

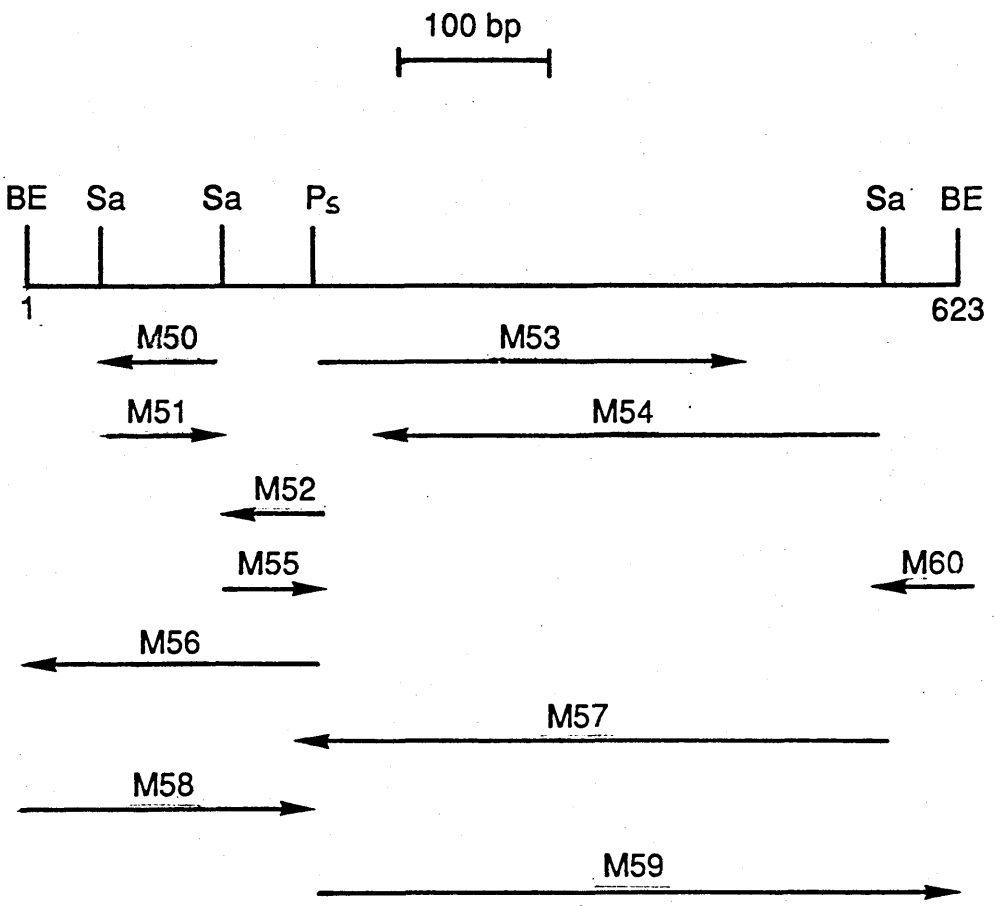
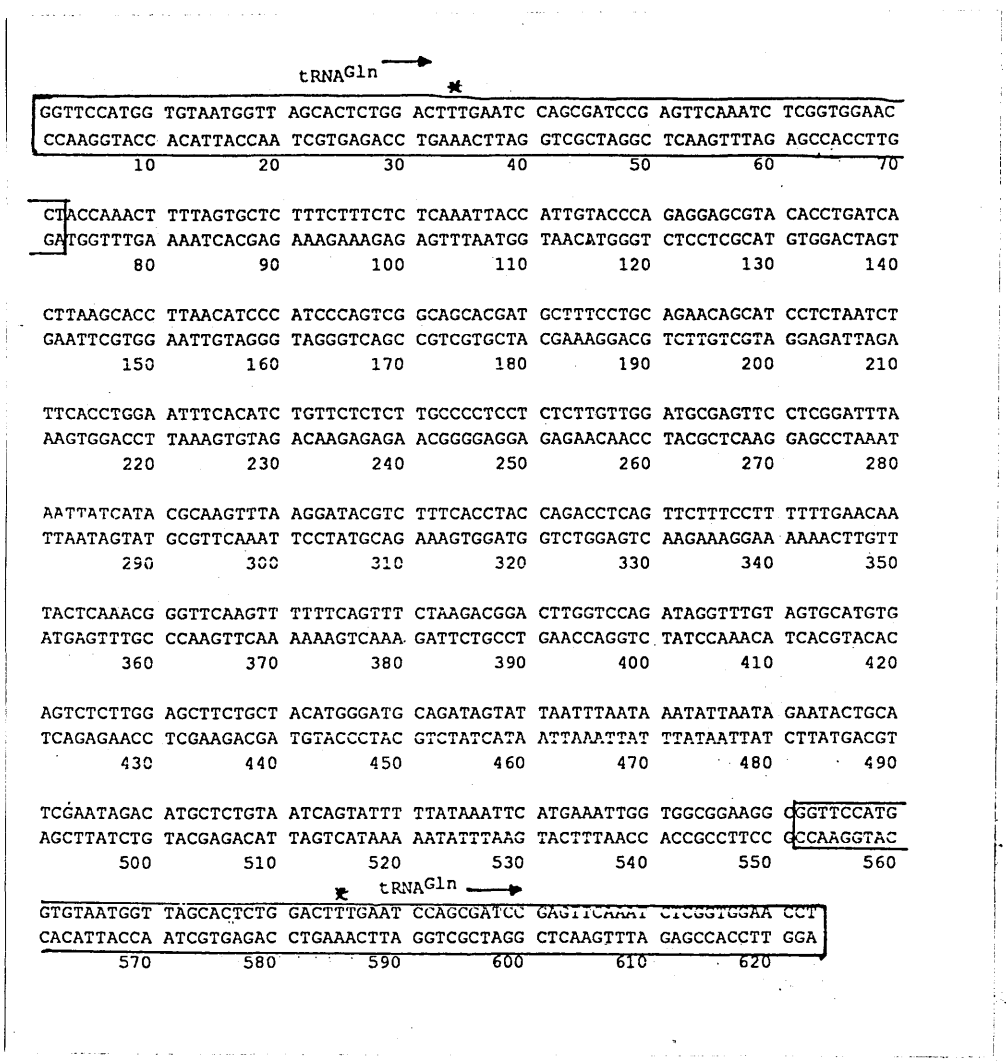


Figure 4.10. Nucleotide sequence of a 623 bp fragment of human genomic DNA containing two tRNA^{Gln} genes.

The nucleotide seqence of a 623 bp fragment of human genomic DNA is shown.The positions, orientation and cloverleaf secondary structure of the tRNA$^{Gln}_{UUG}$ genes are indicated. The sequence was determined for both strands with appropriate overlaps (see figure 4.9.). The difference in gene sequence from the tRNA$^{Gln}_{CUG}$ gene of λHt363 is marked (*).



tRNA genes in question not being alleles showing restriction site polymorphisms, although the number of differences in the restriction maps of λHt363, λHt4 and λHt9 (shown in figure 3.11.) suggests that this is not particularly likely.

Southern hybridisations of genomic DNA can be used to address both of these questions. If no rearrangements have occured in the recombinant then the sizes of genomic restriction fragments hybridising to probes derived from λHt363 can be predicted from the restriction map shown in figure 3.3. Probes that appear to represent sequences common to λHt363 and λHt4 or λHt9 should hybridise to the same genomic restriction fragments. Conversely, probes including sequences outside regions of similarity should detect additional genomic fragments where the restriction maps of the recombinants differ. The genomic DNA used in this study is from a different individual to the source of the DNA used to make the λCharon 4A library. If high frequency restriction site polymorphisms exist between different alleles, and therefore between different individuals, then the pattern of hybridisation obtained using probes derived from λHt363 would not necessarily correspond to that predicted by the restriction map. This follows because the restriction map is derived from the DNA of the original λCharon 4A library.

The genomic hybridisation results presented in sections 4.2.1. and 4.2.2. are consistent with predictions made on the basis of the restriction maps of λHt363, λHt4 and λHt9 in that the probes used detect restriction fragments of the expected sizes (see section 4.2. and figures 4.1. and 4.3.). However, a complex pattern of genomic hybridisation generally occurs with labelled tRNA gene probes, as a consequence of their multigene family nature. This appears to be true for some small restriction fragments containing tRNA genes as well as for gene-specific probes generated by the PCR containing only the tRNA gene coding sequence. It is difficult therefore to draw too many definite conclusions from limited hybridisation data. When a number of bands hybridise to a probe there is a possibility that the "expected" bands may be other genomic fragments hybridising to the probe with a coincidental resemblance in size to fragments predicted on the basis of restriction maps. In the case of this study, a range of genomic digests were used as targets for the hybridisation of both tRNA gene-specific probes and of restriction fragments containing tRNA genes. Fragments of genomic DNA corresponding to the restriction map of λHt363 were detected for each digest, so despite the reservations expressed above the experimental evidence is consistent with the insert of λHt363 representing an unaltered stretch of human genomic DNA.

4.4.2. Copy numbers of the tRNA genes from λHt363.

The discovery that tRNA gene-specific probes produced using the PCR hybridised to multiple genomic fragments (section 4.2.2.) was not unexpected. As described in section 1.4., tRNA genes generally belong to multigene families. If these occur in different parts of the genome then different restriction fragments will hybridise to a tRNA gene probe in a genomic Southern hybridisation. The number of restriction fragments detected in a genomic digest gives an estimate of minimum copy number for the tRNA gene used as a probe. It can be seen in figure 4.3. that the hybridisation of the probe to genomic fragments is not of uniform intensity. This suggests that two or more members of a tRNA gene family may share common genomic environments so that they are located on the same size of restriction fragment for one or more restriction endonucleases. The total number of bands hybridising to a tRNA gene probe therefore represents a minimum estimate of copy number, since each band may be the result of hybridisation to more than one genomic restriction fragment. (The intensity of the bands produced in these experiments was not quantified because of the high background seen on the autoradiographs. This high background, occurring even under stringent hybridisation conditions, may be a consequence of the use of short (71-88 bp) gene specific probes for Southern hybridisations). It should be noted that hybridisation to genomic fragments could be due to probes detecting either tRNA pseudogenes resembling the genes in the cluster, or members of another gene family of isoaccepting tRNA genes of similar sequence to those occurring on λHt363. (Section 4.3.3. describes how two human $tRNA^{Gln}$ genes differing in only one residue from that on λHt363 were cloned). If pseudogenes or different isoacceptor tRNA genes are detected by a gene specific probe then the estimate of minimum copy number for the tRNA genes found on λHt363 would be affected.

As already discussed in section 3.4., the tRNA gene cluster cloned on λHt363 appears to be present in at least two copies in the human genome. The results from the quantification of the copy number of part of the cluster, presented in section 4.2.3., are consistent with this. The actual figure of 2.7 copies per haploid genome is obviously subject to some experimental error, since the true copy number must

be an integer. Likely error sources are in the spectrophotometric quantification of the DNA, in loading samples onto agarose gels, in uneven hybridisation to the nylon membrane and in the densitometry itself since slightly different results can be obtained by scanning different parts of the same band. Two or three copies per haploid genome may be the true figure. Under the conditions used, only one genomic fragment from a *Bam* HI digest hybridised to the 4.2 kb *Bam* HI probe from λHt363. This would be expected on the basis of the restriction map of λHt363, but not necessarily that of λHt4 since only one of the *Bam* HI sites is present on this recombinant. The second *Bam* HI site, if it occurs in the same position in both copies of the cluster, would not be present in the insert of λHt4 (see figure 3.11.). Since only one fragment was detected in a *Bam* HI genomic digest, it appears that the organisation of the two clusters is the same over at least 4.2 kb.

As described in section 4.2.3., no estimate of genomic copy number was obtained for a probe which should detect the $tRNA^{Gly}$ genes from λHt363 and λHt9. It may be that the relatively short length of this probe (356 bp as opposed to the 4.2 kb fragment mentioned above) caused problems with cross-hybridisation to other members of the $tRNA^{Gly}$ gene family.

4.4.3. Chromosomal location of the tRNA gene cluster from λHt363.

Probes derived from λHt363 have been used in a collaborative study to map human tRNA genes to their respective chromosomes. Two regions of λHt363, subcloned into pUC19, were used for this purpose. One subclone (pCLU1; see section 5.4.) contained the 4.2 kb *Bam* HI fragment from λHt363 which carries the genes coding for $tRNA^{Lys}$, $tRNA^{Gln}$, $tRNA^{Leu}$ and $tRNA^{Arg}$. The other subclone contained a 1.3 kb insert which includes the $tRNA^{Gly}$ gene. This insert consists of the sequence from the *Pvu* II site to the 5' of the $tRNA^{Gly}$ gene (see figure 5.9.) to the *Eco* RI site marking the start of the right arm of the λ recombinant (see figure 3.3.). *In situ* hybridisation experiments carried out by Norma Morrison, as described in Morrison *et al* ., (1991) mapped both of these probes to the same site on the short arm of chromosome 17, 17p12-p13.1. (This

localisation has subsequently been refined to 17p13.100-105 using hybridisation of the 4.2 kb probe to somatic cell hybrid DNAs; Dr. D.H. Ledbetter, personal communication). Since both of these probes were isolated from one recombinant it is not unexpected that they both map to the same chromosomal site. However, the resolution of *in situ* hybridisation using light microscopy is only at the megabase level, so this result does not actually prove that the two probes recognise sequences within a few kb of each other in the genome, although the result is obviously consistent with this interpretation. What this does suggest is that all of the copies of this tRNA gene cluster may occur in the same part of the genome since the 4.2 kb probe used is thought to represent DNA common to λHt363 and λHt4, and the 1.3 kb probe DNA common to λHt363 and λHt9. These fragments are thought to be the same or very similar in the different recombinants on the basis of sequence, restriction map and hybridisation data. It was not possible to use probes covering only regions where sequence identity had been established between the recombinants because of the limitations of the *in situ* hybridisation method used. Probes below a certain size will not hybridise sufficiently well to a chromosomal site to give a significant signal, and 1.3 kb is at about the lower limit for successful use in this technique (N. Morrison and E.Boyd, personal communication).

The coding sequence of tRNA genes only covers 70-95 bp. It follows that for most restriction fragments carrying tRNA genes the sequence 5' and 3' of the gene will be larger than the gene itself. The consequences of this for Southern hybridisations were briefly discussed in section 4.2.2. The point made there, that hybridisation will depend more on the nature of the region flanking the gene than on the gene itself, applies at least equally to *in situ* hybridisation experiments. In such experiments, it is the genomic environment of a tRNA gene which is being mapped to a chromosomal locus, rather than an individual tRNA gene. This explains why another group, using tRNA gene probes against DNA from somatic cell hybrids, mapped a human $tRNA^{Gly}$ gene with an identical coding sequence, but different flanking sequences, to chromosome 16 (McBride *et al* ., 1989). In the study described by Morrison *et al* ., the $tRNA^{Gly}$ gene from λHt363 mapped solely to chromosome 17.

4.4.4. Investigations into tRNA gene organisation using the polymerase chain reaction.

The results obtained using tRNA gene probes in genomic Southern hybridisations suggest that the insert of λHt363 has not undergone any rearrangements (see section 4.2.). The results of experiments using the PCR to compare amplified fragments from λHt363 and genomic DNA were consistent with this interpretation. All of the combinations of oligonucleotide primers (designed using sequence data from λHt363) which produced an amplified fragment using λHt363 DNA as a target produced the same size of amplified fragment with a genomic DNA target.

Investigations into the nature of genomic copies of $tRNA^{Arg}$ genes demonstrated the usefulness of the PCR approach for rapidly characterising genomic sequences without having to clone them. The results presented in section 4.3.2. are not entirely unexpected since it has been shown for yeast tRNA genes that the size and sequence of introns are often conserved in different members from within one tRNA gene family, although not between different tRNA genes (Szekely *et al* ., 1988). (This is not true for all intron-containing tRNA gene families; for example, members of a Drosophila $tRNA^{Tyr}$ gene family have introns of different sizes (Suter and Kubli, 1988)). No human $tRNA^{Arg}$ genes have previously been cloned, and no intron-containing $tRNA^{Arg}$ genes have been found in any organism. As mentioned in section 3.4.2., this may reflect problems in detecting intron containing genes in a genomic DNA library rather than their actual occurrence in eukaryotic genomes. The results from this study suggest the existence of a human $tRNA^{Arg}$ gene family with at least seven members, all of which contain an intron of the same size and sequence (see sections 4.2.2. and 4.3.2.). It was mentioned earlier (section 3.4.) that no $tRNA^{Arg}$ gene is recorded on λHt4 despite the fact that it is within the region of similarity between λHt4 and λHt363 (see figure 3.11.). DNA from this region of the insert of λHt4 was kindly supplied by Dr. R. Buckland. When this DNA was used as a target for the PCR, with primers G and H (see table 4.1.) specific for the $tRNA^{Arg}$ gene, a $tRNA^{Arg}$ gene sized fragment was amplified. This suggests that the tRNA gene cluster on λHt4 does include a $tRNA^{Arg}$ gene. amplifications of λHt4 DNA using other primers which correspond to sequences within the cluster indicate that the $tRNA^{Arg}$ gene is in the

same position relative to the other tRNA genes on λHt4 as it is on λHt363. Presumably the $tRNA^{Arg}$ gene was not originally detected on λHt4 for the same reason as it was originally missed on λHt363, i.e. poor hybridisation to the tRNA probes used (see section 3.4.2.).

The use of the PCR to clone novel tRNA genes has recently been discussed (Green *et al* ., 1990). These authors suggest that if two tRNA genes are located together but in opposite orientations then a single PCR primer targeted for either the 5' or the 3' end of the tRNA gene should amplify the tRNA gene dimer. These authors also suggest that if two genes coding for the same tRNA are located together in the same orientation then the PCR amplification of the smaller monomeric gene would be greatly favoured over that of the tRNA gene dimer. The results presented in section 4.3.3. show that this is not entirely true. Even under conditions favourable for the amplification of short DNA fragments a fragment corresponding to a tRNA gene dimer can be amplified sufficiently to be easily detectable. The fragment described, containing two $tRNA^{Gln}$ genes, has not been characterised beyond its sequence. This is largely because no full length fragments could be cloned. (Figure 4.9. shows the subclones which were obtained for sequencing). The most likely reason for this problem is the instability of recombinants containing the full fragment due to the direct repetition of the 72 bp at either end of the fragment.

The small size of tRNA genes and their large copy number in the genome make these genes particularly suitable for cloning using the PCR. This is further discussed in section 6.1.4.

CHAPTER FIVE

IN VITRO TRANSCRIPTION ASSAYS.

5.1. Background and aims.

An homologous *in vitro* transcription assay using a HeLa S3 cell extract is described in section 2.5.16.2. Aliquots of one cell extract preparation (see section 2.5.16.1.) were used for all of the experiments described in this chapter.

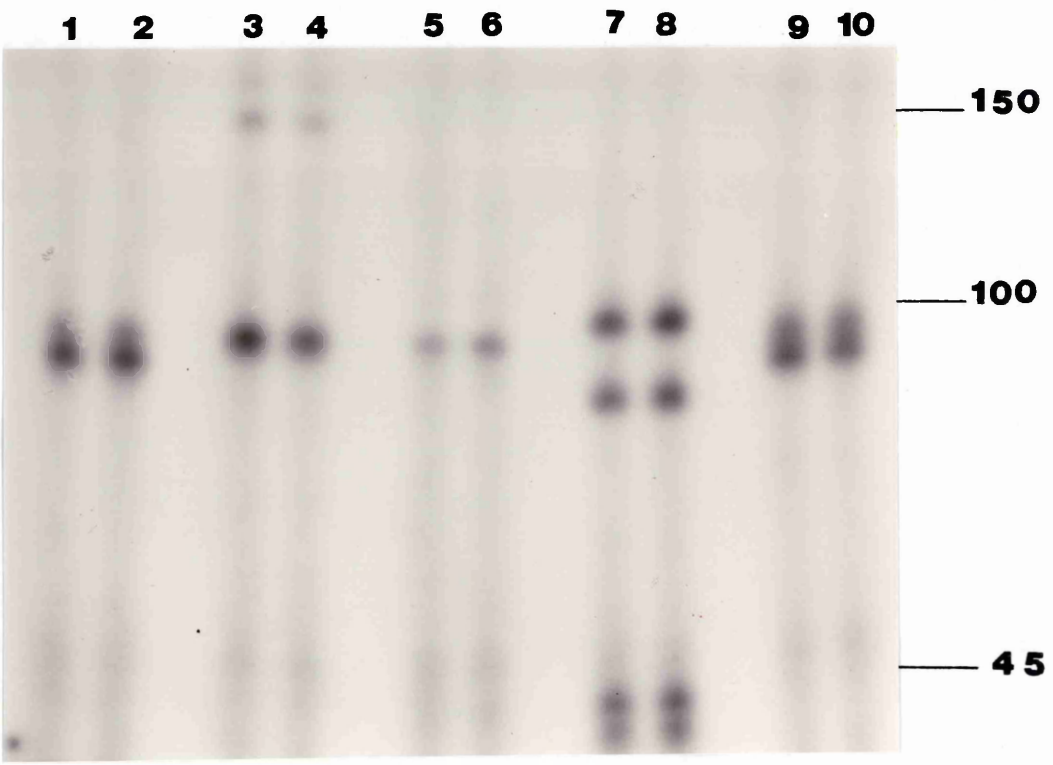
The initial aim of the transcription assays was to establish that each of the tRNA genes from λHt363 was transcriptionally active in the *in vitro* system used, and if so to determine the maximum transcriptional efficiency for each of the genes (in terms of transcripts per gene copy per hour) as described in section 5.3. Establishing this maximum transcriptional efficiency would allow comparison of the rates of transcription of the individual genes to those of the same genes when transcribed as a cluster.

Figure 5.1. shows the gene products from the five tRNA genes on λHt363 separated by polyacrylamide gel electrophoresis.When the gene products from transcription assays are separated by electrophoresis, especially when using sequencing gels (see section 2.5.14.3.2. and figures 5.6. and 5.7.) it is apparent that the majority of the products formed are precursor tRNAs. It was therefore hoped to investigate the processing of the tRNA genes from λHt363, by re-incubation of ^{32}P- labelled gene products in a non-radioactive transcription assay (see section 5.5.). In particular it was hoped to investigate the processing of the $tRNA^{Arg}$ gene, because of the presence of an intron within this gene.

The final group of transcription assays were to investigate the effect of genetic manipulations in the 5' flank of the $tRNA^{Gly}$ gene on its transcriptional efficiency. This gene in particular was selected partly on the basis of having a series of convenient restriction sites in its 5' flank, and therefore being amenable to genetic manipulation. Doran *et al*. (1988) also reported transcriptional differences between a $tRNA^{Gly}$ gene identical to that from λHt363 and a second human $tRNA^{Gly}$ gene

Figure 5.1. Autoradiograph of transcripts of the five tRNA genes from λHt363.

In vitro transcription assay products were electrophoresed in duplicate on an 8% denaturing polyacrylamide gel. Each standard assay (section 2.5.16.2.) contained approximately equimolar amounts of gene template. Lanes 1 and 2, pGLN1 (5.7 ng); lanes 3 and 4, pLEU1 (5 ng); lanes 5 and 6, pLYS1 (5.3 ng) lanes 7 and 8, pARG1 (6.7 ng); lanes 9 and 10, pGLY1 (5.9 ng). (See table 5.1. for a description of the subclones used). Approximate sizes of the products are shown in nucleotides.



with the same coding sequence but different 5' and 3' flanking sequences. This suggests that the flanking regions of these human $tRNA^{Gly}$ genes may influence their transcription, providing a second reason for selecting the $tRNA^{Gly}$ gene from λHt363 for this investigation.

5.2. Construction of recombinants for transcription assays.

All of the quantitative transcription assays performed in this study used supercoiled plasmid DNA as described in section 2.5.16.2. In order to minimise any effect that vector DNA might have in causing transcriptional differences, all of the tRNA genes used were subcloned into the same vector, pUC19 (Yanisch-Perron *et al* .,1985). This vector was also used as non-specific "carrier" DNA in transcription assays (see section 2.5.16.2.). Subcloning was carried out as described in section 2.5.8. The recombinants listed in table 5.1. were those used for all of the transcription assays described in this chapter except for those involving four genes transcribed as a cluster (see section 5.4.), those involving variants of the $tRNA^{Gly}$ gene (see sections 5.6.1. and 5.6.2.) and those involving a variant of the $tRNA^{Gln}$ gene (see section 5.6.4.). The restriction sites used for subcloning can be located on the restriction map of λHt363 (figure 3.3.) and in the sequence data presented (Chapter 3, figures 3.6., 3.8. and 3.10.) and all either occur in the multicloning site of pUC19 or are compatible with enzymes cutting within this site.

5.3. Determination of transcriptional efficiency.

5.3.1. Concentration assays.

When increasing amounts of template DNA are used in transcription assays, the transcriptional apparatus eventually becomes saturated. When the availability of RNA polymerase III and/or transcription factors begins to be limiting, sub-optimal transcriptional efficiency results. For experimental determination of the maximum

Table 5.1.

pUC19 subclones used in single gene transcription assays.

Subclone	Gene	Insert	Sequence reference
pLYS1	$tRNA^{Lys}$	*Bam* HI-*Kpn* I 440 bp	1-440 figure 3.8.
pGLN1	$tRNA^{Gln}$	*Eco* RI-*Eco* RI 617 bp	794-1410 figure 3.8.
pLEU1	$tRNA^{Leu}$	*Eco* RI-*Hin* d III 247 bp	1405-1651 figure 3.8.
pARG1	$tRNA^{Arg}$	*Hin* d III-*Mst* I 1277 bp	1652-2922 figure 3.10.
pGLY1	$tRNA^{Gly}$	*Pvu* II-*Stu* I 713 bp	1-713 figure 3.6.

This table lists the subclones employed in transcription assays performed to establish the maximum transcription rate for the five individual tRNA genes from λHt363. All of the inserts listed were cloned into the multiple cloning site of pUC19 (see figure 2.2. and section 2.5.8.). The numbers in the sequence reference column correspond to the nucleotide residue numbers in the sequence indicated.

transcription rate of tRNA genes it is therefore necessary to use an amount of DNA low enough so that the number of gene copies present in the assay is the factor determining the amount of transcript produced.

To establish the range of DNA amounts low enough to be non-saturating in a transcription assay, concentration curves were produced for each of the five tRNA genes from λHt363. Increasing amounts of plasmid DNA were incubated in a standard transcription assay (see section 2.5.16.2.). The radioactivity incorporated in the transcription products was measured as described in section 2.5.16.3., and plotted against the amount of DNA, and hence the gene copy number, present as template in each assay. Figure 5.2. shows an autoradiograph of transcripts in such a concentration assay performed for the $tRNA^{Leu}$ gene. Similar results were obtained for transcripts from each of the other four tRNA genes from λHt363. (The upper band seen in each lane in figure 5.2. was initially thought to be a precursor of $tRNA^{Leu}$. The exact nature of this approximately 150 nucleotide band is uncertain, but it does not appear to be a RNA polymerase II transcript on the grounds that the band is still produced in the presence of 10 μg/ml α-amanatin. The 245 bp insert of the pLEU1 recombinant (see table 5.1.) does not contain any RNA polymerase III gene other than that for the $tRNA^{Leu}$. This band also appears to be processed in experiments of the type described in section 5.5. and may represent a minority transcript originating from transcriptional read-through of the major transcription termination signal rather than being a $tRNA^{Leu}$ precursor. The sequence of the $tRNA^{Leu}$ gene and its flanking regions is shown in figure 3.8. A putative secondary transcription termination site (T_4) can be seen between residues 1413 and 1416, 87 bp 3' of the end of the coding sequence of the gene. Termination at this site may be responsible for the larger transcript seen in figure 5.2).

The graphs shown in figure 5.3. summarise the results of two duplicate concentration assays for each of the five tRNA genes. Assays using each of the genes gave a similar shape of curve, showing an initial linear increase in cpm with increasing amounts of DNA, tailing off into a plateau as the extract becomes saturated. (If still greater amounts of DNA are added, the total number of transcripts produced falls: E.S. Gonos and J.P. Goddard, personal communication, and data not shown). These results allowed the selection of suitable amounts of DNA from the linear part of the curve to be used in time course assays.

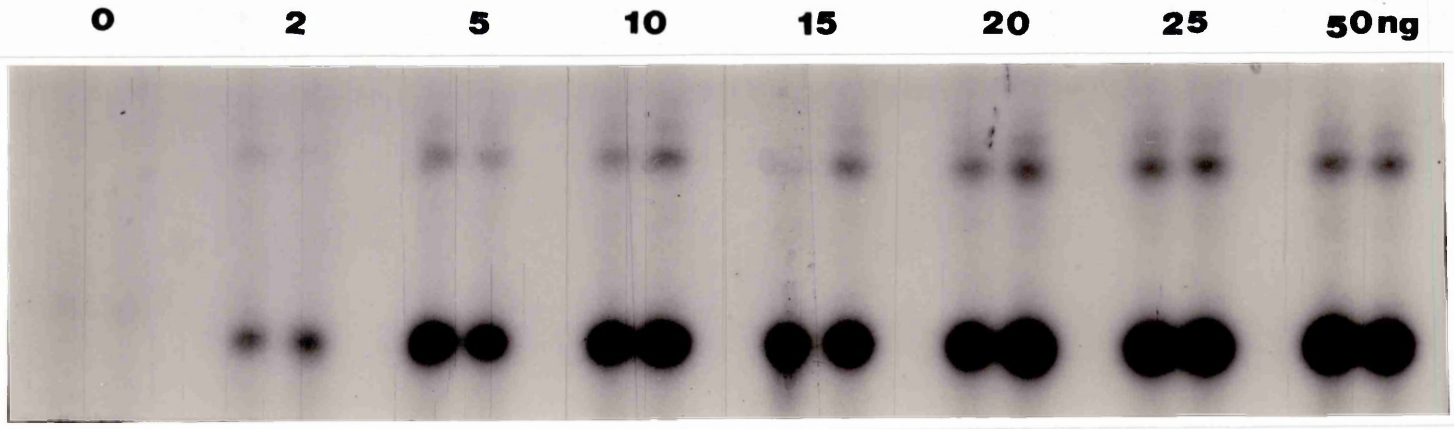
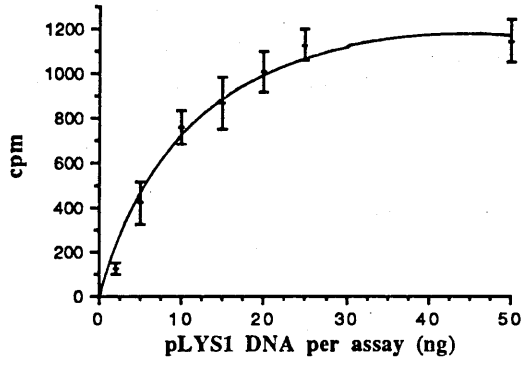


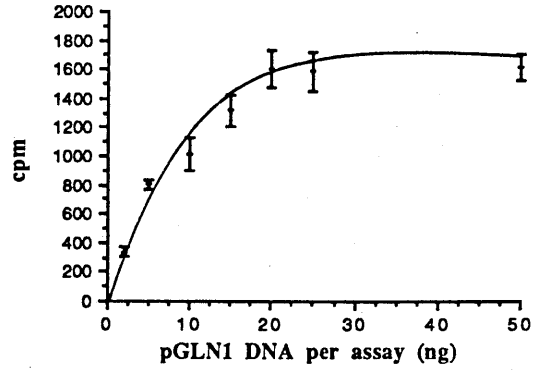
Figure 5.2. Autoradiograph of transcripts from pLEU1: DNA concentration assays.

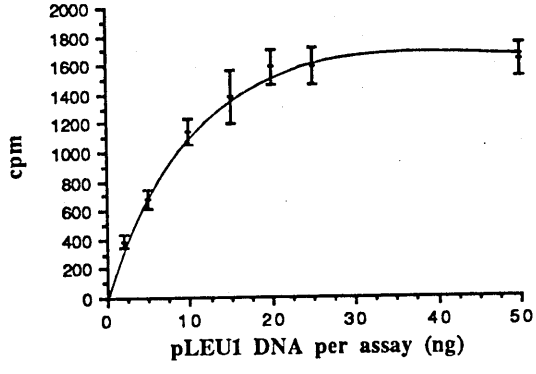
An autoradiograph of products of transcription from 0- 50 ng of pLEU1 (see table 5.1.) is shown. All assays (standard conditions; section 2.5.16.2.) are shown as duplicates.

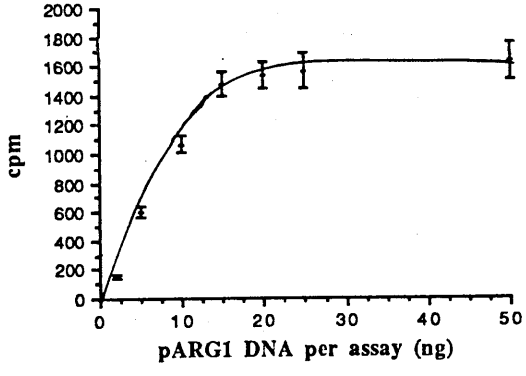
Figure 5.3. Analysis of DNA concentration assays.

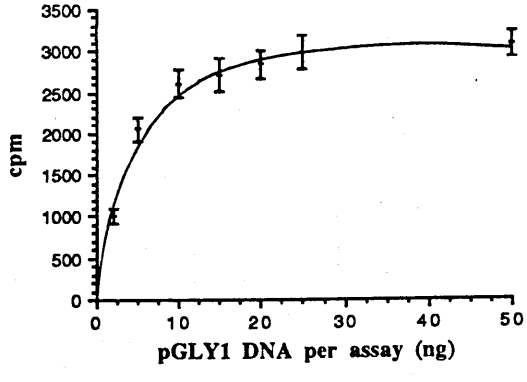
The amounts of [^{32}P] labelled gene transcripts in cpm derived from transcription of 0-50 ng of the indicated subclones are shown (+). Error bars indicate ± one standard deviation for four results (duplicates of each of two independent assays).











5.3.2. Time course assays.

Time course assays were performed for each of the tRNA genes from λHt363 using incubation times between 0 and 120 minutes. All assays used 5 ng of DNA except for those involving the pGLY1 recombinant (see table 5.1.) where 2 ng of plasmid DNA was found to be an appropriate value with reference to the range of concentrations from the linear part of the graphs shown in figure 5.3. Figure 5.4. shows the transcripts generated from the recombinant pLEU1 in transcription assays with incubation times between 0 and 120 minutes. Similar results were obtained for transcripts from the other tRNA genes from λHt363. The amount of transcript was quantified as described in section 2.5.16.3., and plotted as cpm against times of incubation. The graphs in figure 5.5. summarise the results of the time course assays for the five tRNA genes. The graphs all show a similar shape of curve, with an initial lag (due presumably to the time taken for the formation of transcription complexes) followed by a linear increase in labelled transcripts with time. The upper time limit for a continued linear increase in transcripts was not determined. The radioactivity incorporated into transcripts per hour for each gene were determined by taking the difference between cpm recorded at 60 and 120 minutes.

Knowing the amount of transcripts produced by a given number of gene copies per hour, it is possible to calculate the maximum transcriptional efficiency for each gene using the method given in section 2.5.16.4. These figures are shown in table 5.2. for the five genes in terms of transcripts per gene copy per hour. The maximum transcription rates vary by greater than 5-fold from the least efficient to the most efficient gene ($tRNA^{Lys}$ and $tRNA^{Gly}$ respectively).

5.4. Resolution of transcripts on sequencing gels and transcription assays of four tRNA genes as a cluster.

One of the recombinants constructed by subcloning into pUC19 was designated pCLU1, because it carries a 4.2 kb *Bam* HI fragment from λHt363 which had previously been partially characterised and shown to contain a cluster of three tRNA genes coding for $tRNA^{Lys}$, $tRNA^{Gln}$ and $tRNA^{Leu}$ (see section 3.1.).

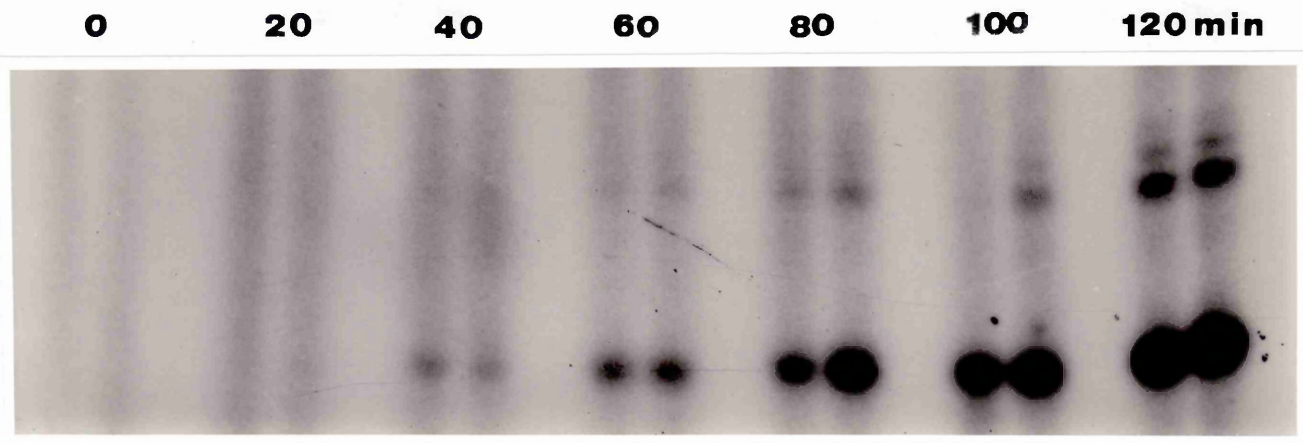


Figure 5.4. Autoradiograph of transcripts from pLEU1: time course assays.

An autoradiograph of products of transcription of pLEU1 (5 ng) for incubation times from 0-120 minutes is shown. All assays (standard conditions; see section 2.5.16.2) are shown as duplicates.

Figure 5.5. Analysis of time course assays.

The amounts of [^{32}P] labelled gene transcripts in cpm derived from transcription of the subclones indicated for 0, 20, 40, 60, 80, 100 and 120 minutes of incubation are shown (+). Error bars indicate ± one standard deviation for four results (duplicates of each of two independent assays).

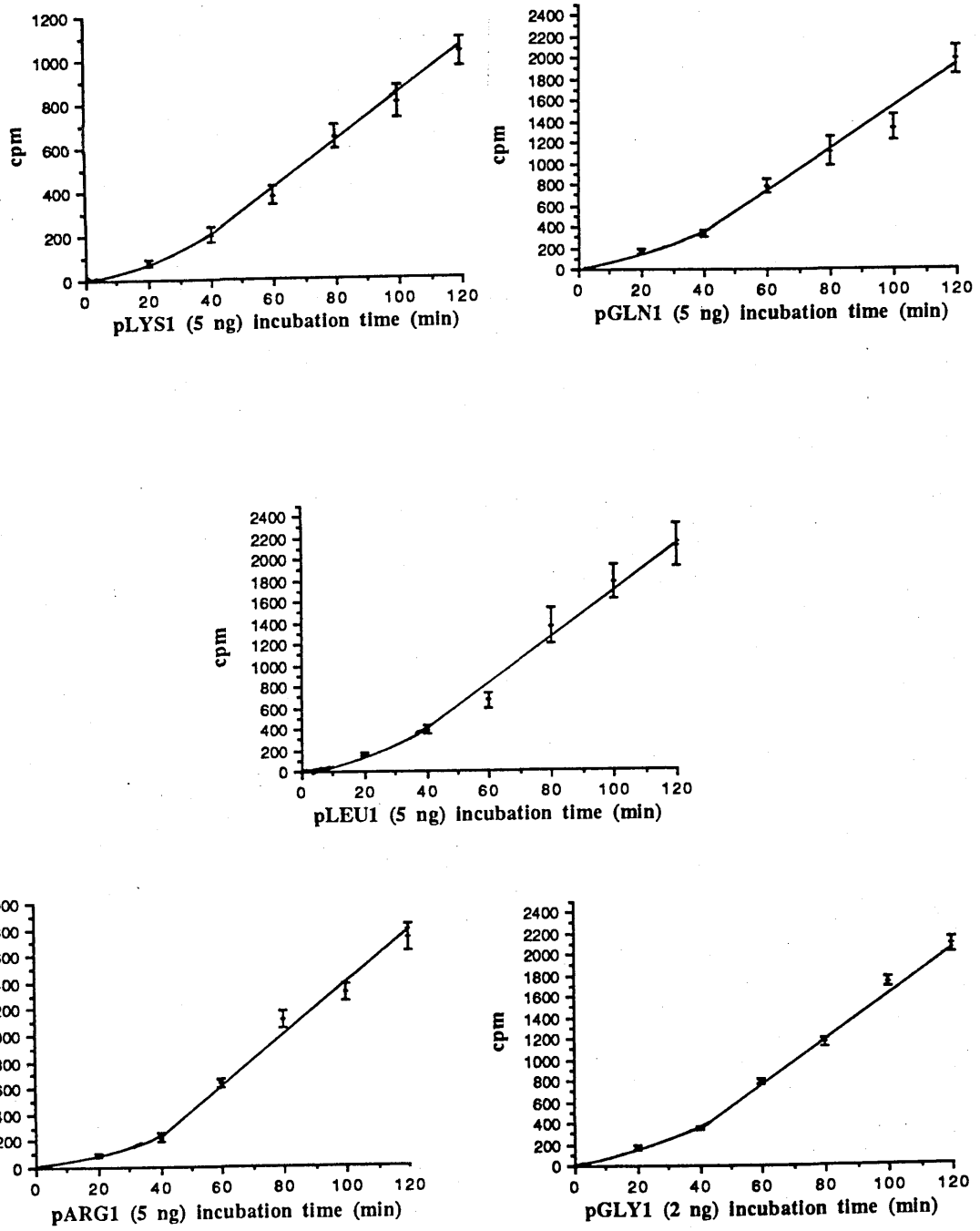


Table 5.2.

Maximum transcription rates for the five tRNA genes from λHt363.

Gene	Transcripts per gene copy per hour	Relative transcriptional efficiency (individual tRNA genes)	Relative transcriptional efficiency (four genes transcribed as a cluster)
$tRNA^{Gly}$	9.0	5.3	–
$tRNA^{Arg}$	3.8	2.2	2.4
$tRNA^{Leu}$	3.5	2.0	2.1
$tRNA^{Gln}$	3.3	1.9	1.7
$tRNA^{Lys}$	1.7	1.0	1.0

This table lists the maximum transcription rates for the five tRNA genes from λHt363 in terms of transcripts per gene copy per hour (see section 5.3.). The recombinants used for the relevant transcription assays were those listed in table 5.1. Table 5.2. also lists the relative transcriptional efficiency for each gene (taking the least active, the $tRNA^{Lys}$ gene as 1.0). Where appropriate, the relative transcriptional efficiencies for the genes when transcribed as a cluster are also tabulated (see section 5.4.).

Figure 5.6. shows an autoradiograph of labelled transcripts derived from pCLU1 separated on a sequencing gel. The autoradiograph also shows transcripts from the recombinants pLYS1, pGLN1 and pLEU1 (see table 5.1.) which contain the three individual genes known at this stage to be carried on pCLU1, and from the recombinant pGLY1 which contains the $tRNA^{Gly}$ gene from λHt363. It can be seen from this autoradiograph that some transcripts are obtained from pCLU1 which do not correspond in size to transcripts from the three genes known to be present on this recombinant. As described in section 3.3.3., the region responsible for producing these bands was subcloned and sequenced and found to contain a $tRNA^{Arg}$ gene which accounted for the additional transcripts observed. It can also be seen in figure 5.6. that the tRNA precursors produced by *in vitro* transcription of the tRNA genes from λHt363 can be resolved on a sequencing gel.

Figure 5.7. shows an autoradiograph from a sequencing gel on which transcripts from pCLU1 have been electrophoresed in parallel with those from the recombinants (described in table 5.1.) containing the five individual tRNA genes from λHt363.The different transcripts from pCLU1 are sufficiently resolved to make it possible to quantify the transcripts from each individual gene. In table 5.2. the ratios of transcriptional efficiencies of the five individual tRNA genes from λHt363 are shown in addition to the corresponding ratios from the four tRNA genes found on pCLU1. Within the limits of this experimental system it appears that these genes are transcribed with similar relative efficiency in a cluster as they are as individual genes despite their close proximity.

5.5. Processing of tRNA precursors.

Figures 5.6. and 5.7. show that the transcripts derived from the tRNA genes used in this study are made up of what appears to be a mixture of precursor tRNAs. One approach to investigate the processing of these putative precursors is to take transcripts from a standard transcription assay as described in section 2.5.16.2. and re-incubate in the HeLa S3 extract in the absence of α-[^{32}P]-UTP so that new transcripts are not labelled, and any size change in the labelled transcripts with time can be monitored. The sizes of transcripts were determined by running the products of a standard sequencing reaction (section 2.5.14.3.1.) alongside the tRNA gene products on a sequencing gel as shown in figure 5.6. In particular it was decided to

Figure 5.6. Preliminary characterisation of transcripts from subcloned tRNA genes.

Transcripts were obtained using standard assay conditions from the $tRNA^{Gly}$, $tRNA^{Leu}$, $tRNA^{Gln}$ and $tRNA^{Lys}$ genes on recombinants pGLY1, pLEU1, pGLN1 and pLYS1 (see table 5.1.). These transcripts were electrophoresed on a sequencing gel in parallel with transcripts from the recombinant pCLU1, a pUC19 recombinant containing a 4.2 kb *Bam* HI fragment from λHt363 (see section 5.4.) which at this stage of the study was thought to contain only the $tRNA^{Lys}$, $tRNA^{Gln}$ and $tRNA^{Leu}$ genes. Lanes 1 and 10, pGLY1 (2 ng); lanes 2 and 9, pLEU1 (5 ng); lanes 3 and 8, pGLN1 (5 ng); lanes 4 and 7, pLYS1 (5 ng) and lanes 5 and 6, pCLU1 (5 ng).The products of a standard sequencing reaction for the M13mp18 subclone M1 (see table 3.2. and figure 3.4.) were run in parallel (tracks A,G,C and T) to estimate the approximate size of transcripts, indicated in nucleotides.

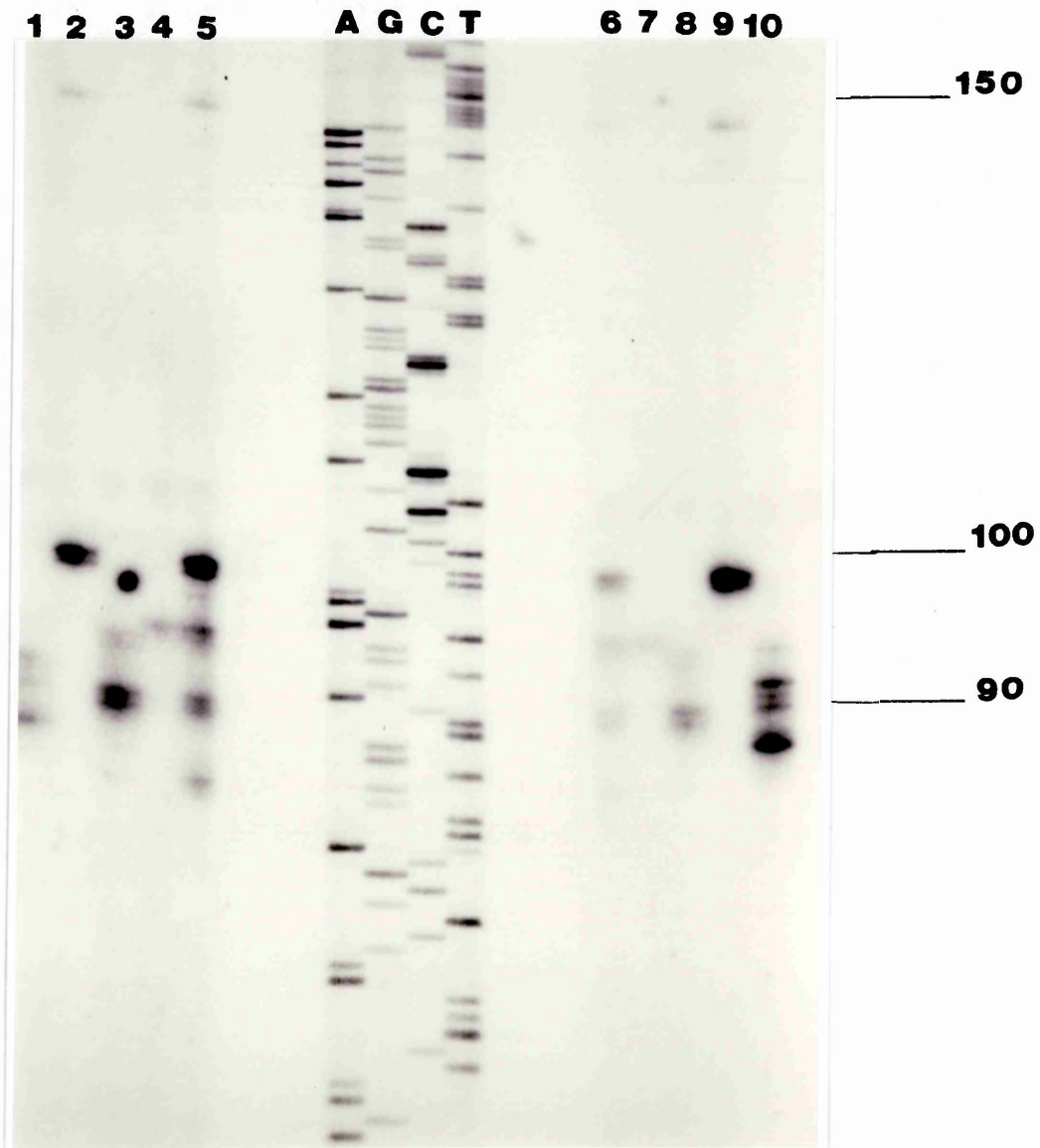
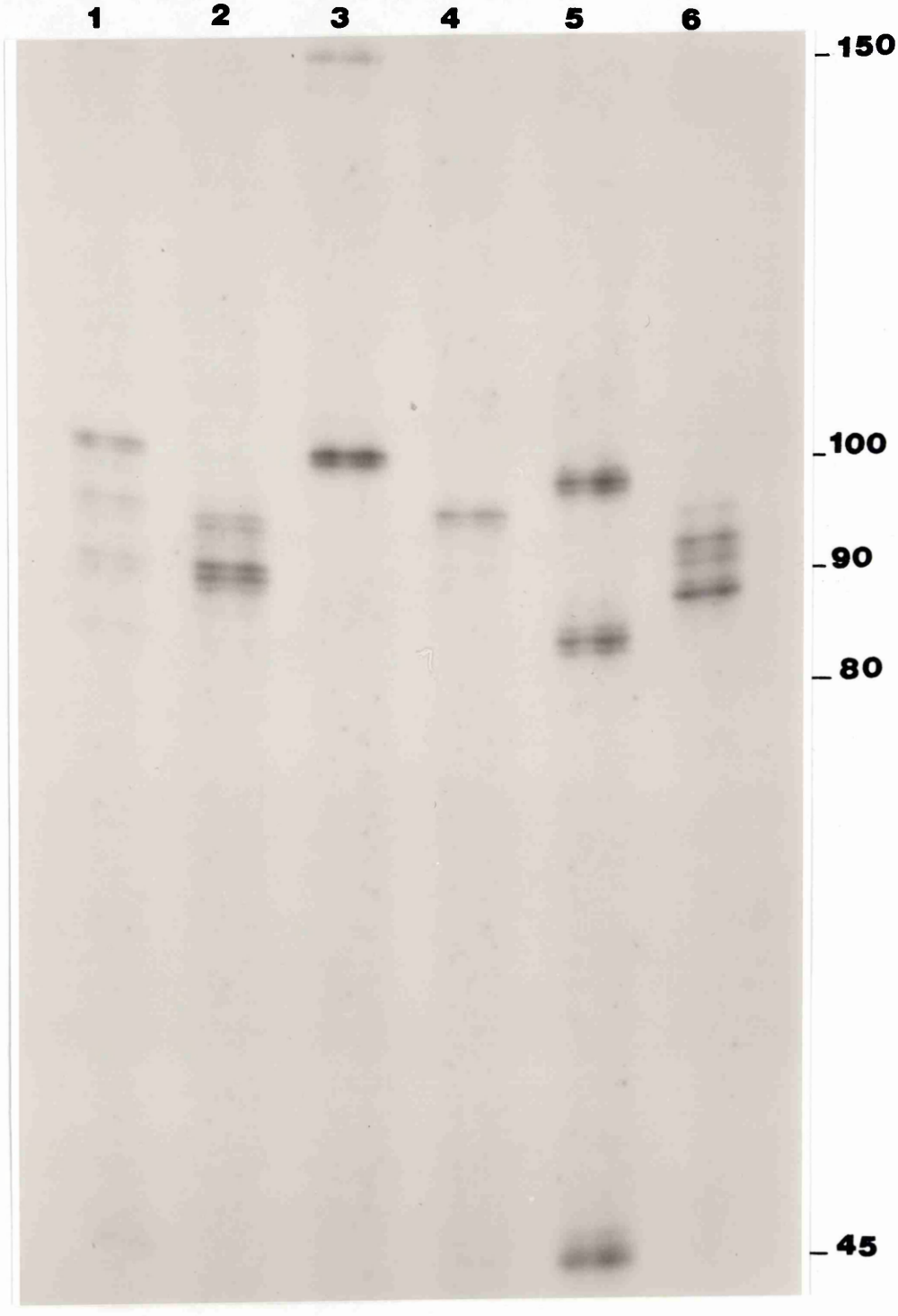


Figure 5.7. Autoradiograph of the products of transcription from the five tRNA genes subcloned from λHt363.

The transcription products of four of the tRNA genes from λHt363, subcloned as a cluster (pCLU1, see section 5.4.) and of the five individually subcloned genes (pGLN1, pLEU1, pLYS1, pARG1 and pGLY1; see table 5.1) were obtained using standard assay conditions, using, in the case of the individually subcloned genes, approximately equimolar amounts of gene template. Lane 1, pCLU1 (5 ng); lane 2, pGLN1 (5.7 ng); lane 3, pLEU1 (5 ng); lane 4, pLYS1 (5.3 ng); lane 5, pARG1 (6.7 ng) and lane 6, pGLY1 (5.9 ng). Products were separated on a sequencing gel; their approximate sizes are shown in nucleotides.



look at the changes in size of the precursors for $tRNA^{Arg}$, since the removal of the intron and splicing of the resultant half molecules causes the most obvious differences in precursor size.

5.5.1. Processing of $tRNA^{Arg}$ gene products.

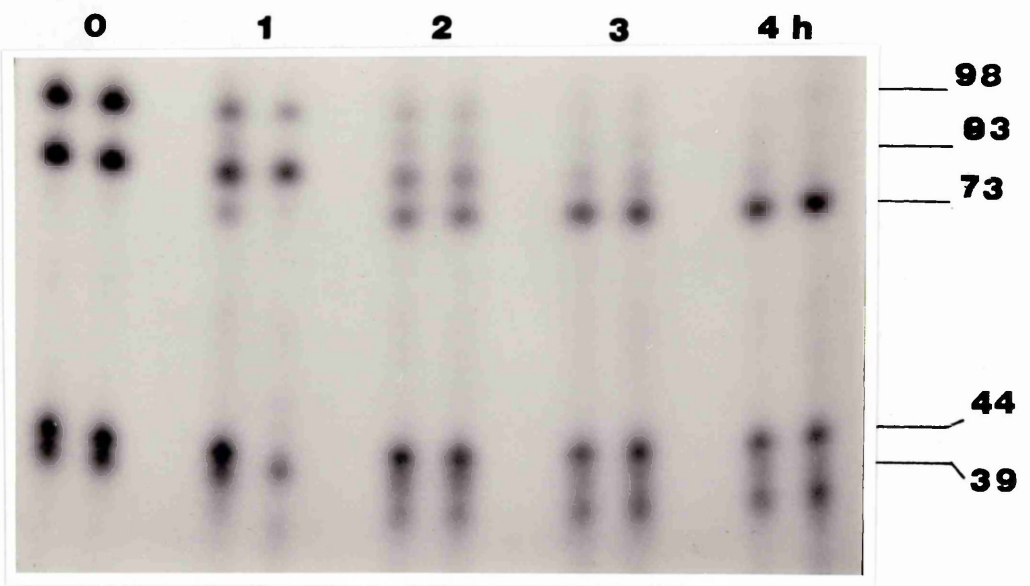
Figure 5.8. shows the results of re-incubating the products of a one hour transcription assay using pARG1 (see table 5.1.) for times between 0 and four hours. The sizes of the products are indicated in figure 5.8. There is an accumulation with time of a product of the size of $tRNA^{Arg}$, and the expected loss of the largest precursor corresponding in size to the tRNA coding sequence plus intron plus 10 nucleotides of flanking sequence. A similar pattern of changes was observed when the pARG1 recombinant was used in time-course assays of the type described in section 5.3.2. Re-incubation experiments for the other tRNA genes in the cluster gave similar results, with the eventual accumulation of tRNA-sized products after prolonged re-incubation (data not shown).

The approximate sizes of the bands seen in figure 5.8. were 98, 83, 73, 44 and 39 residues, together with some faint intermediate bands and indeterminate products of less than 39 residues which are visible after two hours of re-incubation. These sizes and the order in which the $tRNA^{Arg}$ precursors appear are consistent with the following scheme of processing. Firstly, the intron is removed from the primary transcript of 98 bases to yield two tRNA half molecules of 44 and 39 bases, and presumably release the 15 base intron although no intron-sized band was resolved on the gel (see figure 5.8.). The two half molecules are spliced to produce the 83 base product which still has the 5' leader and 3' extension sequences. These are apparently removed together since no intermediate products can be seen between the 83 base band and the tRNA-sized product of 73 residues. It is uncertain whether the addition of the CCA tail found at the 3' terminus of mature tRNAs occurs in the *in vitro* system used. O'Connor and Peebles (1991) report that splicing of tRNA half molecules can take place either before or after 5' and 3' end trimming. The presence of products of less than 39 residues could be explained by the occurence of either 5' or 3' end maturation before splicing of the tRNA half molecules.

The total length of the 5' leader plus 3' trailer in the primary $tRNA^{Arg}$

Figure 5.8. Autoradiograph of the products of processing of precursor tRNAs derived from pARG1.

The products of a transcription assay using pARG1 (10 ng) were reincubated in a HeLa cell extract in the absence of [^{32}P]UTP for 0-4 h. Duplicate samples were removed at the times indicated and separated by electrophoresis through an 8% denaturing polyacrylamide gel. Approximate product sizes are indicated in nucleotides.



transcript was 10 residues (assuming that the CCA tail had not been added to this 98 residue transcript). This suggests that there are only four potential transcription start sites for the $tRNA^{Arg}$ gene, the purine residues (G in each case) at -7, -4, -3 or -2 (see figure 3.10.). The residue at -7 may be the best candidate for the start site since the termination signal for the $tRNA^{Arg}$ gene occurs one bp 3' of the end of the coding sequence, so the 3' extension is unlikely to extend much beyond this point. Assuming that this is the case the 44 base product corresponds to the 5' tRNA half molecule and the 39 base product the 3' half molecule.

5.6. The effect of the 5' flank of the $tRNA^{Gly}$ gene on transcription.

5.6.1. Construction and transcription of variant $tRNA^{Gly}$ gene recombinants.

Variants of the pGLY1 recombinant (table 5.1.) were constructed using restriction sites in the 5' flank of the $tRNA^{Gly}$ gene. Figure 5.9. lists the recombinants constructed, with differing amounts of the 5' flank removed. pGLY3 differs from the other constructs in that part of the 5' flank (125 bp bounded by the two *Bam* HI sites denoted by B^1 and B^2 in figure 5.9.) has been placed in the opposite orientation to that found in λHt363. The *Mst* I site used in the construction of pGLY5 cuts exactly flush with the start of the coding sequence of the gene, so in this recombinant the entire 5' flank is replaced with unrelated vector DNA. Transcription assays were performed, using the same amount of DNA for each variant, and the efficiency of transcription of the variants pGLY2, 3, 4 and 5 relative to pGLY1 calculated. Figure 5.10. shows an autoradiograph of the products of a transcription assay using these 5' variants. The percentage efficiencies (the mean of duplicates from two independent assays) were as follows: pGLY1 = 100%, pGLY2 = 102%, pGLY3 = 51%, pGLY4 = 83% and pGLY5 = 78%. As well as differences in the amounts of transcript, variants appeared to produce slightly different patterns of precursors (see figure 5.10.). It has previously been reported that changes in the 5' flank of tRNA genes can lead to the utilisation of alternative sites for the start of transcription and hence different sizes of tRNA precursor (Hofstetter *et al*., 1981; Hipskind and Clarkson, 1983).

Figure 5.9. Construction of pUC19 recombinants containing $tRNA^{Gly}$ gene variants.

The restriction sites indicated (*Bam* HI (B^1 and B^2), *Mst* I (M) *Pvu* II (P) and *Stu* I (St)) were used to construct the pUC19 recombinants pGLY1-pGLY5, with the variations shown in the 5' flank of the $tRNA^{Gly}$ gene. The inversion of the *Bam* HI fragment B^1-B^2 in pGLY3 is indicated (←). The $tRNA^{Gly}$ gene itself is represented by a solid box.

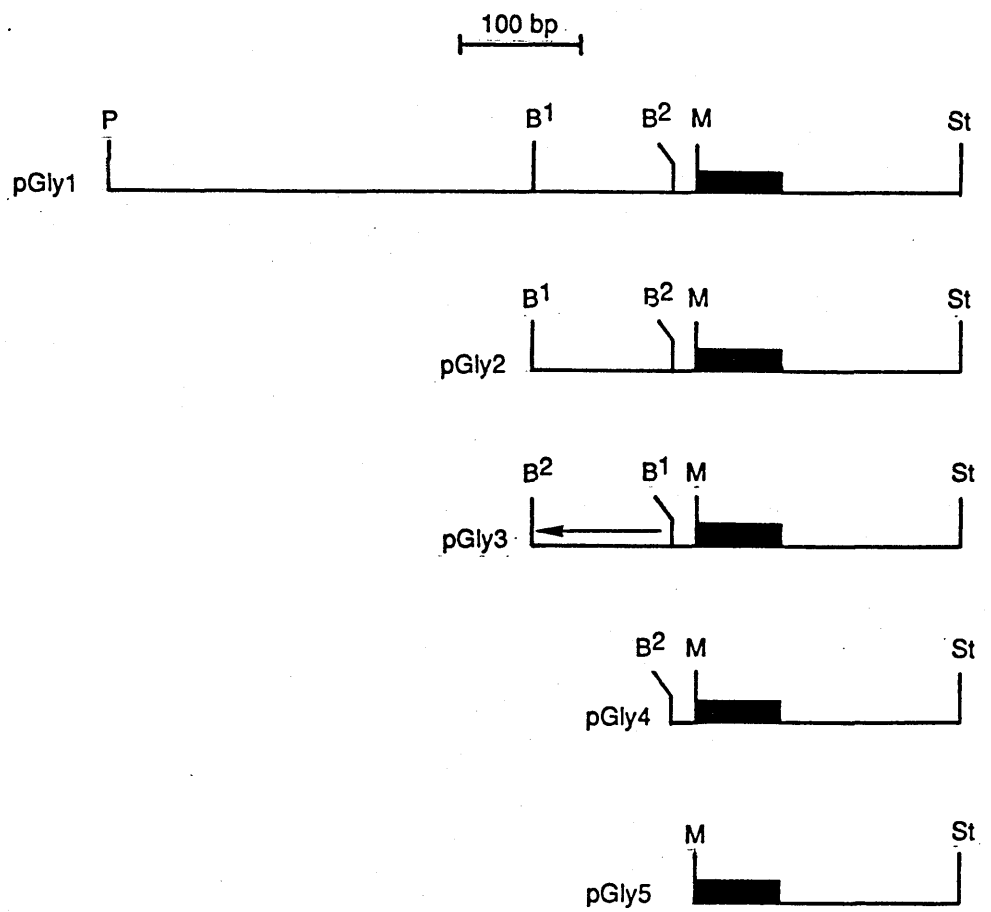
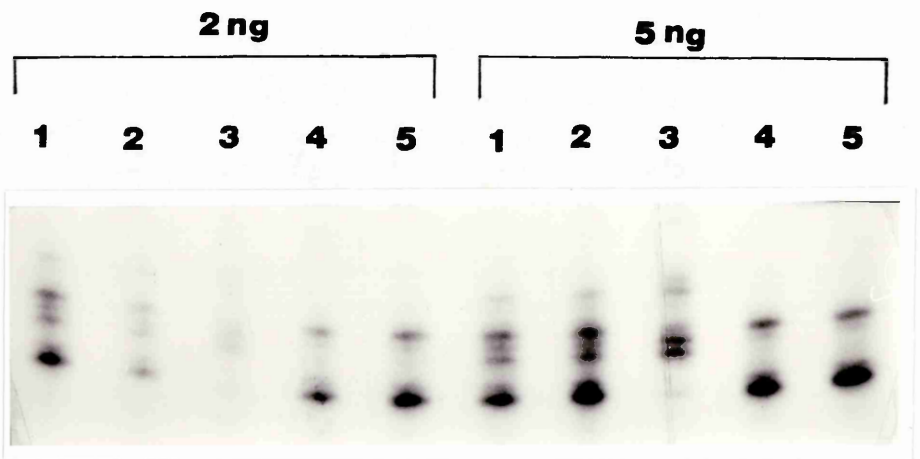


Figure 5.10. Autoradiograph of the transcription products from $tRNA^{Gly}$ gene variants.

The recombinants pGLY1-pGLY5 (see figure 5.9.) were used in *in vitro* transcription assays and the products analysed on a sequencing gel. The products for two different amounts of each recombinant (2 ng and 5 ng) are shown. Lanes 1-5 correspond to recombinants pGLY1-pGLY5.



These results largely agree with previous reports, referred to in section 1.5.6., in that deleting DNA from the 5' flank of the gene and replacing it with unrelated vector DNA causes a slight reduction in transcription. This effect is not very marked, since in pGLY5 the entire 5' flank is replaced causing only a 22% decrease in transcripts. However, a more marked decrease in transcription was observed for pGLY3, in which part of the flanking sequence has been placed in the opposite orientation to the original. This raises the possibility that the 125 bp *Bam* HI fragment involved may play a role in transcription complex formation.

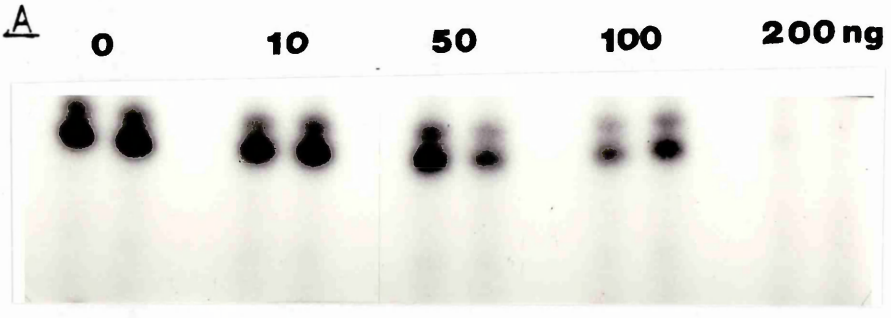
5.6.2. Competition assays

Since the 125 bp *Bam* HI fragment from the 5' flank of the $tRNA^{Gly}$ gene appeared to have an effect on transcription, a recombinant was constructed by subcloning this fragment into the *Bam* HI site of pUC19. This recombinant, termed pGLY5'F as it contained part of the $tRNA^{Gly}$ gene 5' flank, was used as a competitor in a series of transcription assays. Increasing amounts of the pGLY5'F recombinant (0-200 ng) were added to transcription assays containing 5 ng of pGLY1. (Adding equivalent amounts of pUC19 had previously been shown to have no effect on transcription). An autoradiograph of the transcription products from such assays is shown in figure 5.11.A. These products were quantified and the results are summarised in the graph shown in figure 5.11.B. The addition of pGLY5'F to the transcription assays does appear to interfere with the transcription of the $tRNA^{Gly}$ gene on pGLY1. The addition of 200 ng of pGLY5'F, representing an approximately 40-fold molar excess over the $tRNA^{Gly}$ gene, was sufficient to reduce transcription to less than 10% of the original level. In subsequent experiments, the pGLY5'F construct was used as a competitor with the other tRNA genes from λHt363. Very similar results were obtained, implying that the fragment of $tRNA^{Gly}$ gene 5' flanking sequence subcloned in pGLY5'F may be able to sequester some general tRNA transcription factor present in the HeLa cell extract.

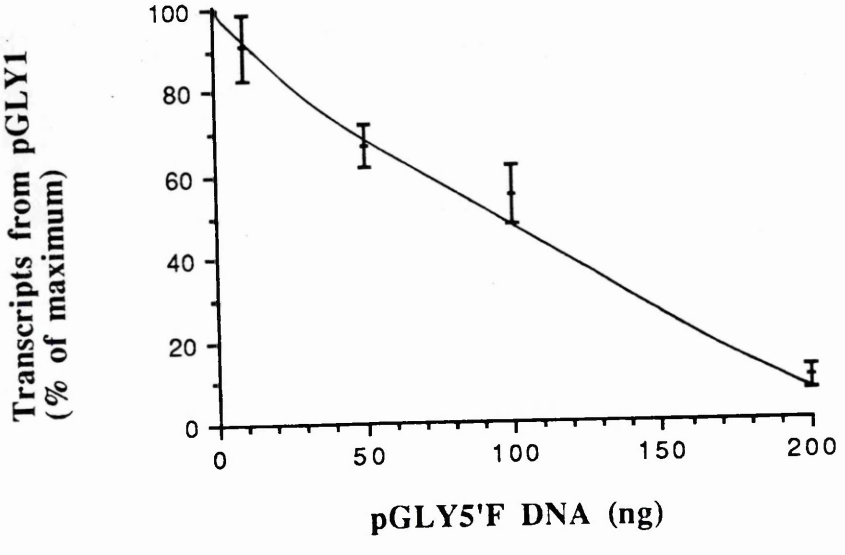
Figure 5.11. Autoradiograph of products of transcription of pGLY1 in the presence of increasing amounts of pGLY5'F and analysis of competition assays.

A Standard transcription assays of pGLY1 (5 ng) were performed in the presence of 0-200 ng pGLY5'F competitor DNA (see section 5.6.2.). The transcription products, shown in duplicate, were separated on an 8% polyacrylamide gel.
B The amounts of [^{32}P] labelled gene transcripts in cpm derived from transcription of pGLY1 (5 ng) with 0, 10, 50, 100 and 200 ng of pGLY5'F competitor DNA are shown (+). The error bars indicate ± one standard deviation for four results (duplicates of each of two independent assays)

A



B



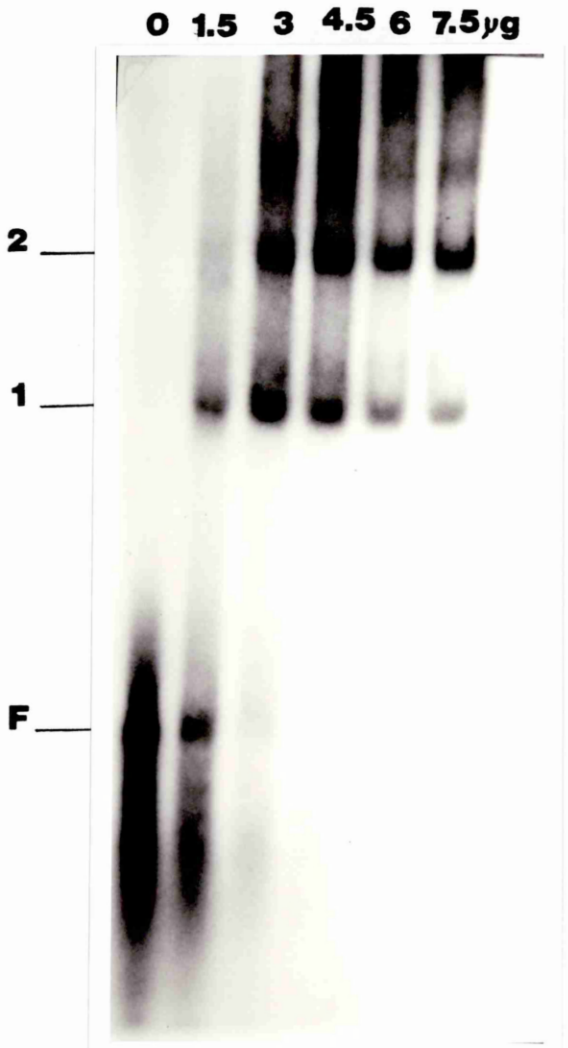
5.6.3. Gel-shift and DNase I protection assays.

The results shown in section 5.6.2. suggest that the 125 bp *Bam* HI fragment from the 5' flank of λHt363 may be able to associate with some factor or factors present in the crude HeLa cell extract used in the transcription assays. Two 19-mer oligonucleotides were designed in order to specifically amplify the 175 bp region immediately 5' of the $tRNA^{Gly}$ gene using the polymerase chain reaction (see section 2.5.15.1.). The oligonucleotides correspond to residues 328-346 of the upper (tRNA-like) strand and 502-484 of the lower strand of the nucleotide sequence shown in figure 3.6. and hence can be used to amplify a region which includes the 125 bp *Bam* HI fragment from the $tRNA^{Gly}$ gene 5' flank. This 175 bp amplified fragment was end-labelled at the 5' position of one strand with γ-[^{32}P]-ATP as described in section 2.5.11.3. and incubated with the HeLa cell extract prior to use in gel shift and DNase I protection assays (see sections 2.5.17. and 2.5.18.). Figure 5.12. shows that a mobility shift does occur after incubation of the end-labelled fragment in the crude extract. As well as the free labelled DNA, two lower mobility bands can be seen on the autoradiograph. The intensity of these retarded bands increases with increasing protein concentration, and figure 5.12. also shows that the faster migrating of the two putative protein/DNA complexes appears after incubation with a lower concentration of protein from the crude extract. This smaller band also appears after a shorter incubation time with a given amount of extract (data not shown).

This pattern of bands would be consistent with that obtained from the sequential binding of two different proteins (or two molecules of the same protein) to the end-labelled restriction fragment. To assess the specificity of the interaction causing the gel shift, competition assays were performed. The competitor DNAs used were a specific competitor (the amplified 175 bp fragment from the 5' flank of the $tRNA^{Gly}$ gene, unlabelled and used at up to 40-fold molar excess) and a non-specific competitor (a 141 bp *Sau* 3A fragment of pUC19, again unlabelled and used at up to 40-fold molar excess). Unexpectedly, the non-specific competitor DNA served to displace the end-labelled 175 bp amplified fragment from the protein/DNA complexes almost as well as the specific competitor at a 20- and 40-fold molar excess (data not shown). This finding would argue against a strong, specific association of the fragment from the $tRNA^{Gly}$ gene 5' flank with a putative transcription factor or factors.

Figure 5.12. Autoradiograph from a gel shift assay of protein interaction with part of the $tRNA^{Gly}$ gene 5' flank.

Increasing amounts of a HeLa cell extract (0, 1.5, 3, 4.5, 6 and 7.5 µg of protein) were incubated with a 175 bp PCR amplified fragment of $tRNA^{Gly}$ gene 5' flanking DNA (see section 5.6.3.) end-labelled with ^{32}P. The resulting gel shift pattern was analysed by electrophoresis on a 10% polyacrylamide gel. The positions of the free DNA (F) and of the putative protein-DNA complexes (1 and 2) are indicated.



In order to further investigate the possible binding of proteins in the HeLa cell extract to the 175 bp amplified fragment, DNase I protection (footprinting) assays were carried out as described in section 2.5.18. The pattern of DNase I digestion appeared to be identical for the free and the bound fractions of the 5' end-labelled 175 bp fragment as shown in figure 5.13. for DNA labelled on the complementary strand of the $tRNA^{Gly}$ gene 5' flank. The same was true for DNA labelled at the 5' end of the opposite strand. Again this would suggest that the protein/DNA interaction may not be a strong one, since the binding of the protein(s) does not appear to give any significant DNase I protection to the DNA fragment.

5.6.4. The effect of the 5' flank of the $tRNA^{Gly}$ gene on transcription of a second tRNA gene.

The recombinant pGLN2 was constructed by taking a 528 bp *Pst* I-*Mst* I fragment from the recombinant pGLY1, comprising 502 bp of the $tRNA^{Gly}$ gene 5' flank and 26 bp of vector sequence, and subcloning it into pGLN1 (see table 5.1.) digested with *Pst* I and *Sma* I, using the *Pst* I and *Sma* I sites from the multicloning site of pUC19. The net effect of these manipulations was to place the complete immediate 5' flank of the $tRNA^{Gly}$ gene, corresponding to residues 1-502 in figure 3.6., 182 bp upstream of the coding sequence of the $tRNA^{Gln}$ gene from pGLN1. The transcriptional efficiencies of pGLN1 and pGLN2 were then compared by using a range of DNA concentrations for the two recombinants in a series of transcription assays (see figure 5.14.). The transcription products derived from the two recombinants were quantified as described in section 2.5.16.3. and these results indicated that there was no significant difference between the transcriptional efficiencies of pGLN1 and pGLN2.

5.7. Discussion.

5.7.1. The transcriptional efficiency of the tRNA genes from λHt363.

The *in vitro* transcription assays used in this study give reproducible and

Figure 5.13. Autoradiograph from a DNase I protection assay.

A 175 bp PCR amplified fragment of $tRNA^{Gly}$ gene 5' flanking DNA, end-labelled with ^{32}P, was subjected to gel shift analysis after incubation with 1.5 µg of HeLa cell extract protein (see figure 5.12.). The bound (B) and free (F) fractions were eluted from a polyacrylamide gel and treated (as described in section 2.5.18.) with two different amounts of DNase I (1 µg and 2 µg) and the resulting DNA fragments resolved on a sequencing gel.

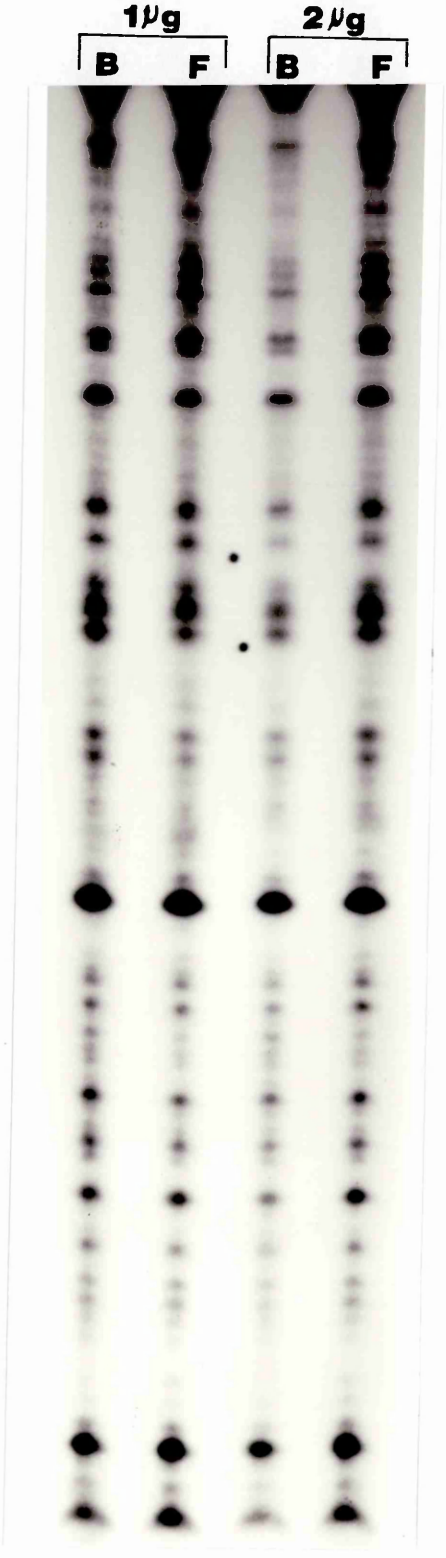
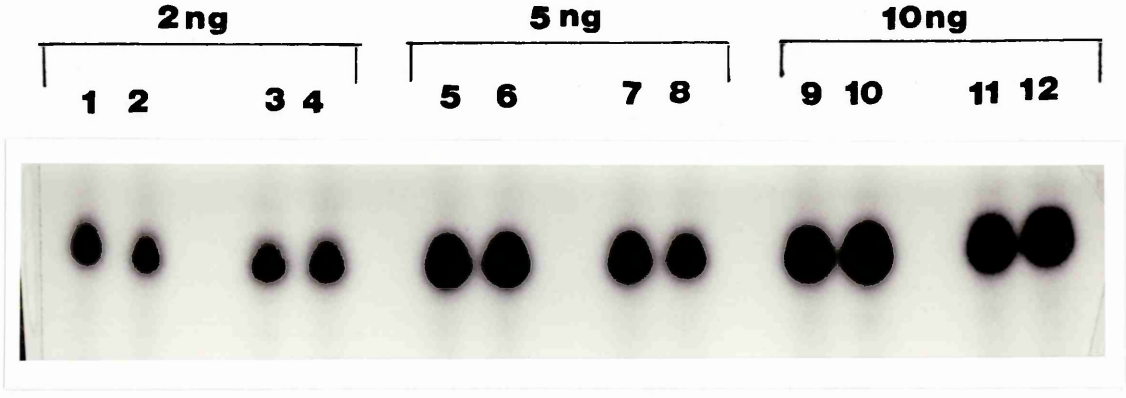


Figure 5.14. Autoradiograph of transcripts from pGLN1 and pGLN2.

Varying amounts, as shown, of recombinants pGLN1 and pGLN2 (each containing the $tRNA^{Gln}$ gene from λHt363 but differing in 5' flanking sequence; see section 5.6.4.) were used as templates in standard duplicate transcription assays and the products electrophoresed on an 8% polyacrylamide gel. Lanes 1,2,5,6,9 and 10 = pGLN1 products, lanes 3,4,7,8,11 and 12 = pGLN2 products.



internally consistent results for each of the tRNA genes characterised from λHt363, agreeing well with previous experiments using the same system (Gonos, 1989; Gonos and Goddard, 1990c). However, this does not necessarily mean that the results accurately reflect the way that the genes are transcribed *in vivo* . It has been shown that tRNA genes are transcribed differently using different cell extracts (Raymond and Johnson, 1983; Johnson *et al* ., 1984; Schaack and Soll, 1985; Drabkin and RajBhandry, 1985a and1985b; Gonos and Goddard, 1990c) and it is thought that transcription may be regulated by species-, tissue- or developmental stage-specific factors which may not be present in any given cell extract. It may therefore be dangerous to overstate the significance of *in vitro* transcription results. A considerable degree of difference was recorded in the maximum transcriptional efficiencies of the five tRNA genes from λHt363, although all of the genes yielded more than one transcript per gene in a one hour period, indicating that multiple rounds of transcription were occurring. Table 5.2. shows that a greater than five-fold difference was found in the numbers of transcripts per gene per hour between the least active gene in this system, coding for $tRNA^{Lys}$, and the most active, coding for $tRNA^{Gly}$. The high activity of the $tRNA^{Gly}$ gene in this system was interesting since Doran *et al* ., (1988) had reported a low transcriptional activity for an identical human $tRNA^{Gly}$ gene, with the same flanking sequences. It is noticable that the order of transcription rates (table 5.2.) follows the order of the tRNA genes on λHt363 from left to right (see figure 3.3.). Whether or not this is of any significance is doubtful. In any case the $tRNA^{Leu}$ gene (figures 3.3. and 3.8.) is transcribed on the opposite strand and therefore in the opposite direction from the other genes in the cluster.

Russo *et al* ., (1987b) reported differences in the transcription of tRNA genes when their positions within a cluster were altered. No such experiments have been attempted for the genes from λHt363. The relative transcription rates for the four most closely spaced genes on the cluster remained the same when the genes were transcribed individually (section 5.4.). However, in the recombinants used the genes still retained at least 100 bp of 5' and 3' flanking sequence. The fact that the same relative transcription rates were maintained may reflect the fact that the genes still have their original control regions, rather than being unaffected by their immediate genomic environment. It is possible that the differences seen in the transcriptional efficiencies of the tRNA genes from λHt363 reflect differences in their internal promoter regions and not in the flanking sequences of the genes. This

would explain why four of the tRNA genes maintained the same relative transcription rates when transcribed either separately or together from a single recombinant. The box B regions of the five genes are compared below with the consensus sequence for this part of the internal promoter. No comparison was attempted for the box A element since this region is less tightly conserved (see section 1.5.2.). Nucleotides in the box B elements of the tRNA genes from λHt363 which differ from the consensus sequence are shown in bold type and underlined.

Box B consensus:	GGTTCGANTCC
λHt363 $tRNA^{Lys}$ box B:	GGTTC**A**AGTCC
λHt363 $tRNA^{Gln}$ boxB:	**A**GTTC**A**AATC**T**
λHt363 $tRNA^{Leu}$ box B:	GGTTCGAATCC
λHt363 $tRNA^{Arg}$ box B:	GGTTCGAATCC
λHt363 $tRNA^{Gly}$ box B:	GGTTCGATTCC

The box B regions of the $tRNA^{Lys}$ and $tRNA^{Gln}$ genes differ from the box B consensus sequence in one and three nucleotides respectively. These are the two tRNA genes from λHt363 with the lowest transcriptional efficiencies (see table 5.2.) but there is probably no significant correlation of box B sequence with observed transcription rate. The $tRNA^{Leu}$, $tRNA^{Arg}$ and $tRNA^{Gly}$ genes have identical box B sequences but as table 5.2. shows the $tRNA^{Gly}$ gene is transcribed with greater than two-fold efficiency when compared to either of the other genes. In contrast, the box B regions of the $tRNA^{Leu}$ and $tRNA^{Gln}$ genes differ in three positions but the genes have a very similar transcriptional efficiency in the *in vitro* transcription system used.

The spacing of boxes A and B in the $tRNA^{Leu}$ and $tRNA^{Arg}$ genes from the λHt363 cluster (42 and 48 residues respectively) is larger than that found in the other tRNA genes because of the presence of a large variable arm in $tRNA^{Leu}$ and an intron in the $tRNA^{Arg}$ gene. However, these distances fall within the optimum range for transcription complex formation (34-53 nucleotides; see section 1.5.3.).

5.7.2. Processing of tRNA precursors.

The processing of tRNA precursors in a HeLa cell extract may not reflect the *in*

vivo situation for similar general reasons to those discussed with reference to transcription in section 5.7.1. It has been shown that the optimum conditions for transcription may not be optimum for processing, and that optimum conditions for processing may differ in different systems (van Tol *et al* .,1987). The conditions used in the *in vitro* transcription system described in section 2.5.16.2. do not appear to be conducive to efficient processing. No significant accumulation of mature tRNA-sized products was observed for any of the tRNA genes from λHt363 with less than two hours re-incubation.

The $tRNA^{Arg}$ gene products from the *in vitro* transcription system did appear to be processed in that the sizes of discrete bands changed with increasing times of re-incubation in the HeLa cell extract (section 5.5.1.). The size changes observed were consistent with the pattern of processing described in section 5.5.1. although it should be stressed that no direct confirmation of the sequence of any of the transcripts generated in this study has been made. It is therefore not possible to say with certainty what is represented by any particular precursor tRNA. Hegg and Thurlow (1990) have shown that tRNAs may have residual secondary structures in denaturing 8M urea/TBE polyacrylamide gels, leading to the presence of variant minor bands. The sizes of tRNAs and tRNA precursors were determined in this study by comparison to single-stranded DNA fragments generated in a standard sequencing reaction (see section 2.5.14.3.). If the tRNAs and pre-tRNAs do maintain some secondary structure in the denaturing gels used then inaccuracies in size determination may have occurred.

5.7.3. The role of the 5' flank of the $tRNA^{Gly}$ gene in the control of transcription.

Transcription assays using constructs with changes in the flank of the $tRNA^{Gly}$ gene gave one potentially interesting result. The construct pGLY2 (table 5.1.) had a significantly higher transcription rate than construct pGLY3 which has the same amount of 5' flanking DNA. The only difference between the constructs is that a 125 bp *Bam* HI fragment (spanning residues -21 to -145 relative to the $tRNA^{Gly}$ coding sequence) has been "turned around" in pGLY3 to the opposite orientation to that found in the original genomic DNA. This change has a greater

effect on transcription than the removal of the entire 5' flank of the gene and its replacement with unrelated DNA (see section 5.6.1.). It is possible that the region -21 to -145 of the 5' flank of this human $tRNA^{Gly}$ gene contains an element or elements involved in transcription complex formation or stability, and that the effect of these putative element(s) is dependent either on orientation relative to the coding sequence of the gene or on distance upstream of the gene.

Preliminary experiments to further investigate potential control elements in the 5' flank of the $tRNA^{Gly}$ gene are described in sections 5.6.2., 5.6.3. and 5.6.4. If the 125 bp *Bam* HI fragment reversed in pGLY3 is capable of binding to any transcription factors in the HeLa cell extract then it should be capable of competing for these factors with other tRNA genes. The results presented in figure 5.11.B show the effect of adding different amounts of pGLY5'F DNA to a transcription assay with a fixed amount (5 ng) of pGLY1 DNA (see section 5.6.2.). pGLY5'F does appear to interfere with the transcription of the $tRNA^{Gly}$ gene on pGLY1, but only when in a 20-fold molar excess does the reduction equal that found in pGLY3. pGLY5'F appears to compete with the other tRNA genes from λHt363 to the same extent as it competes with the $tRNA^{Gly}$ gene (data not shown). The effect of pre-incubating pGLY5'F in the HeLa cell extract prior to adding the gene to be competed was tested, but the pre-incubation step did not appear to affect the results of the competition assays (data not shown).

These experiments suggest that if a control element is carried on the 125 bp *Bam* HI fragment from the $tRNA^{Gly}$ gene 5' flank then firstly its effect is much greater in *cis* and secondly its effect in *trans* does not appear to be gene-specific. The $tRNA^{Gly}$ 5' flank was placed upstream of, and therefore in *cis* with, the $tRNA^{Gln}$ gene in the recombinant pGLN2 described in section 5.6.4., with no apparent effect on transcription. However, since the $tRNA^{Gln}$ gene in pGLN2 retains 163 bp of its immediate 5' flank it is likely that any control element placed 5' of this region would be too far upstream to affect the transcription of the gene.

Crude gel-shift and footprinting assays were used to attempt to establish whether the 175 bp amplified fragment from the 5' flank of the $tRNA^{Gly}$ gene does associate with a protein or proteins in the HeLa cell extract (Section 5.6.3.). These experiments gave apparently conflicting results. Figure 5.12. shows that a mobility shift did occur when the end-labelled fragment was incubated in the HeLa cell extract. Despite this apparent evidence for protein binding, the autoradiograph in figure 5.13. shows that the bound fragment did not appear to be protected from

DNase I digestion. A range of DNase I concentrations were used in protection assays but none gave evidence of any footprint. It may be that a protein/DNA association can cause a mobility shift without significantly protecting the DNA from DNase I, possibly because either the association is weak or the protein binds in such a way that still allows DNase I access to the DNA .

The major problem in proposing a model which can explain the differences observed in the transcription of the $tRNA^{Gly}$ gene 5' flank variants lies in reconciling results which appear to be mutually contradictory. Experimental evidence indicates that part of the 5' flank can reduce transcriptional efficiency when inverted relative to the gene, and when used in competition assays, but can be entirely replaced by unrelated DNA without significantly reducing transcription. There is evidence that the 175 bp fragment from the 5' flank of the tRNA gene does bind to a protein or proteins in the HeLa cell extract, but this binding is insufficient to give any DNase I protection. A model consistent with the available data is given below.

The results presented in sections 5.6.2. and 5.6.3. would suggest that there may be an element (or elements) within the region between 21 and 145 bp upstream of the $tRNA^{Gly}$ gene capable of binding a factor (or factors) which can influence either the rate of transcription complex formation or the stability of the transcription complex. The effect of this putative element depends on either its position relative to the coding sequence of the gene or its orientation relative to the gene (or a combination of both). The fact that the $tRNA^{Gly}$ gene is the most transcriptionally active of all the genes tested in this study may reflect the presence of a positively-modulating element in its 5' flank. If such an element exists, this would explain the ability of the recombinant pGLY5'F (which contains the relevant part of the 5' flank of the $tRNA^{Gly}$ gene) to compete with other tRNA genes in a transcription assay. When added in a large enough excess, pGLY5'F may be able to sequester a factor present in the HeLa cell extract which contributes to transcription, possibly interfering with nascent transcription complexes when used to compete with complete tRNA genes. If this putative factor is absolutely essential for transcription it must be capable of binding alternative sequences since the entire 5' flank of the $tRNA^{Gly}$ gene can be replaced with unrelated DNA without a large fall in transcription rate. The "loose" protein association with the 175 bp fragment from the 5' flank of the $tRNA^{Gly}$ gene suggested by preliminary footprinting experiments may be due to the putative factor being unable to bind tightly to the

isolated fragment in the absence of any transcription complex.

CHAPTER SIX

DISCUSSION.

In this chapter an overall analysis of the results presented in Chapters three, four and five will be given and suggestions made for future work to address questions raised by these results. The genomic organisation and *in vitro* expression of the tRNA genes investigated will be discussed in the light of current work in these areas.

6.1. Genomic organisation of the human tRNA genes from λHt363.

6.1.1. Nucleotide sequence analysis of the tRNA gene cluster from λHt363.

Sequence analysis of part of the insert of the recombinant bacteriophage λHt363 revealed the presence of five human tRNA genes within a 7 kb region, representing the largest human tRNA gene cluster so far described. Four of the genes have identical coding sequences to human tRNA genes described previously (Roy *et al* ., 1982; Doran *et al* ., 1988) and as discussed in sections 3.4. and 4.4. it appears that all five genes from λHt363 represent second copies of tRNA genes, of which four have been previously characterised in different genomic environments and the fifth is almost certainly present in a characterised cluster. The occurrence of tRNA genes in small clusters (defining a cluster in this sense as two or more genes within a few kb of each other) appears to be a common feature in eukaryotic genomes (see section 1.4.). As mentioned in section 3.4.2., the fact that a number of such clusters have been cloned may reflect the relative ease of their detection by hybridisation to a tRNA probe in comparison to dispersed genes, rather than being a real indication of the way in which most eukaryotic tRNA genes are organised. As yet, no particular pattern of distribution has emerged for tRNA genes in the genomes of higher eukaryotes.

The nucleotide sequence data obtained in this study (Chapter three and Chapter

four) was analysed using the program ANALYSEQ (Staden, 1986; see section 2.5.14.). In addition to a tRNA gene search, the sequence data from λHt363 was examined for restriction endonuclease recognition sites, open reading frames, RNA polymerase III gene conserved sequences, direct and inverted repeats and finally for the presence of consensus sequences characteristic of *Alu* repetetive elements (Kariya *et al* ., 1987) which have been found in close proximity to several human tRNA genes (Chang *et al* ., 1986; Arnold *et al* ., 1986; Doran *et al* ., 1987; Shortridge *et al* ., 1989; Craig *et al* ., 1989). Apart from the tRNA genes, no sequence motifs of especial interest were discovered, but parts of the insert did prove to be very G+C rich (see section 3.4.3.). The fact that some of the tRNA genes from λHt363, in particular the $tRNA^{Gly}$ gene, are in such a G+C rich environment may have no significance for their expression. However, a recent report (Besser *et al* ., 1990) suggests that the methylation status of CpG dinucleotides in a chicken $tRNA^{Lys}$ gene (either within the coding sequence or in the 5' flank of the gene) can affect its transcription. The effect of methylation of CpG rich islands on the transcription of eukaryotic RNA polymerase II genes is well documented (for review see Bird, 1986). The methylation status of genomic DNA can be investigated either by genomic sequencing (Church and Gilbert, 1984; Saluz and Jost, 1989) or by comparison of the patterns of hybridisation of a probe to genomic DNA cut with one or other of a pair of restriction enzymes with the same recognition site but different sensitivity to methylated nucleotides within that site (Bird, 1978; Bird and Southern, 1978). It may be appropriate to apply these methods to the genomic DNA represented in the insert of λHt363 to investigate whether or not CpG dinucleotides in this region are methylated *in vivo* . The sequence data is available to construct oligonucleotides for use in genomic sequencing, and suitable restriction enzyme sites could be selected for the hybridisation approach. The effect of DNA methylation on *in vitro* transcription could also be investigated.

6.1.2. Comparison of the tRNA genes from λHt363 with members of the same human tRNA gene families cloned on different genomic fragments.

The results presented in Chapters three and four show that the tRNA genes

present on λHt363 are apparently second copies of tRNA genes which had previously been characterised. A comparison of the restriction maps of the recombinants involved, λHt4 and λHt9 (Buckland *et al* ., 1983; Doran *et al* ., 1988), with that of λHt363 is discussed in section 3.4.1. and an investigation into the possibility of rearrangements having occurred within the insert of λHt363 is described in Chapter four. The overall conclusions from these results were that the insert of λHt363 represents a different region of genomic DNA from those present in the inserts of λHt4 and λHt9, and that no apparent rearrangement had occurred. It was also shown in a collaborative study (Morrison *et al* ., 1991; see section 4.4.3.) that probes from the tRNA gene cluster found on λHt363 could be located solely to one site on the short arm of chromosome 17 (17p12-p13.1) by *in situ* hybridisation. This chromosomal assignment has been confirmed and refined (using hybridisation of a probe from λHt363 to a somatic cell hybrid panel) to 17p13.100-17p13.105 (Dr. D. H. Ledbetter, personal communication).This result raises the possibility that different copies of this cluster of five tRNA genes are themselves clustered in the genome. A small tRNA gene cluster in the rat is contained within a 13.5 kb repeated unit which is thought to be arranged as ten tandem repeats (Sekiya *et al* ., 1981; Shibuya *et al* ., 1982 and 1985) and a human $tRNA^{Asn}$ gene family is thought to be located on large (180 kb) repeat units, organised in tandem arrays (Buckland, 1989). The localisation of the tRNA gene cluster from λHt363 to a single chromosomal site would be consistent with a genomic arrangement as tandem repeats of the two to three copies of the cluster. This could be investigated using hybridisations of probes from the tRNA gene cluster to large genomic DNA fragments generated by restriction endonucleases with few sites in human genomic DNA. The restriction endonuclease *Not* I, which has a unique site in the insert of λHt363 might be particularly suitable for this purpose since, as figure 4.1. shows, a probe containing a tRNA gene from the cluster hybridises to high molecular weight fragments in a *Bgl* II/*Not* I genomic digest. Pulse field gel electrophoresis would be the most suitable method of resolving large fragments of genomic DNA. If a tandem arrangement of the cluster does exist then it should be possible to use the hybridisation of a probe derived from within the tRNA gene cluster to partial digests of genomic DNA to demonstrate that the hybridising bands consist of multiples of the repeat unit size.

As described in section 3.4.1., the restriction maps of λHt363, λHt4 and λHt9 differ outwith the areas which contain the tRNA genes (see figure 3.11.). By a combination of cross-hybridisation experiments, fine restriction mapping and finally

nucleotide sequence determination it should be possible to determine the nature of the DNA at the boundaries of the regions of similarity. The sequences found in these regions might give some idea as to the mechanism behind the evolution of different copies of the tRNA gene cluster. If the different copies arose through duplication of an ancestral cluster the extent of the accumulation of neutral base changes in the regions surrounding the tRNA genes would give an indication of the evolutionary timing of the duplication event. It has been suggested that many human tRNA genes arose by RNA-mediated transposition events and represent functional retrogenes (McBride *et al* ., 1989). If the tRNA genes found on λHt363 were retroposons then it would be expected that target site reduplications would be found to flank the tRNA genes.

The *in situ* hybridisation work referred to above (Morrison *et al* ., 1991) localised probes containing tRNA genes from λHt363 to a single chromosomal site. This technique could be used to map other restriction fragments derived from the inserts of λHt363, λHt4 and λHt9 which either come from wholly outside the regions of similarity in the restriction maps or span the areas covering divergence points. If any of these probes proved to hybridise to chromosomal sites apart from 17p13.100-17p13.105 it would suggest that the organisation of the tRNA gene cluster may be more complex than a tandem repeat.

6.1.3. Copy numbers and organisation of the tRNA gene families represented on λHt363.

Each of the tRNA genes found on λHt363 appears to belong to a multicopy gene family with members occurring in different genomic environments (see sections 4.2.1. and 4.2.2.). This pattern, or lack of pattern of organisation seems to be the general rule for human tRNA genes in common with other eukaryotes; most characterised human tRNA gene families have copy numbers of the order of 10-20 per haploid genome (Arnold *et al* ., 1986; Krupp *et al* ., 1988; Zasloff and Santos, 1980; Gonos and Goddard, 1990a). The part of the tRNA cluster found on both λHt363 and λHt4 containing the $tRNA^{Lys}$, $tRNA^{Gln}$, $tRNA^{Leu}$ and $tRNA^{Arg}$ genes has a copy number of 2-3 per haploid genome as determined by genomic hybridisations (results shown in section 4.2.3.). The copy number of the genomic

fragment containing the $tRNA^{Gly}$ gene common to λHt363 and λHt9 was not determined. The problem of cross-hybridisation of the probe from λHt363, presumably to other members of the $tRNA^{Gly}$ gene family were described in sections 4.2.3. and 4.4.2. If such cross-hybridisation occurs it may be an indication of homologies in the flanking regions of different members of the $tRNA^{Gly}$ gene family.

The $tRNA^{Arg}$ gene from λHt363 and its gene family are of particular interest since no $tRNA^{Arg}$ gene from any source and only two other human tRNA gene families had previously been shown to possess introns. It should be noted that as discussed in section 3.4.2. this tRNA gene hybridises poorly to a tRNA probe, suggesting that the lack of characterised tRNA genes with introns may be due in part to technical difficulties in their cloning rather than reflecting their actual representation in the genome. It is interesting that of the three human tRNA gene families shown to have introns, two were characterised using methods that did not directly involve hybridisation (this study and Green *et al* ., 1990) and the third, for $tRNA^{Tyr}$, was isolated using a *Xenopus laevis* $tRNA^{Tyr}$ gene probe which itself contained an intron (MacPherson and Roy, 1986).

The presence of an intron in the $tRNA^{Arg}$ gene from λHt363 did not appear to affect its *in vitro* transcription and the precursor tRNAs produced by the gene were processed to a mature tRNA sized product (see sections 5.3. and 5.5.1.). The role of introns in tRNA genes is uncertain except for the specific case of the intron of $tRNA^{Tyr}$ genes (see section 1.5.5.). It has been suggested that the lack of introns in most higher eukaryote tRNA genes could be due to the origin of multi-gene families by RNA-mediated transposition. Those introns still present in higher eukaryote tRNA genes could represent evolutionary relics (McBride *et al* ., 1989). However, there is no specific evidence to support this speculation.

6.1.4. Applications of the PCR in characterising human tRNA genes.

The sequences of genomic copies of the $tRNA^{Arg}$ gene family were determined as described in section 4.3.2. and the sequencing artifacts from templates generated by asymmetric PCR (figure 4.7.) were also briefly mentioned. The pattern of "artifact" bands outside the coding region of the gene is not uniform; as can be seen in

figure 4.7. regions of apparently readable sequence are interspersed with regions of strong compressions. It is uncertain whether these observations have any real significance. The "artifact" sequence seen outside the coding region did not appear to correspond to any of the flanking regions of the $tRNA^{Arg}$ gene from λHt363 but as figure 4.7. shows the sequence data obtained from this part of the autoradiograph is ambiguous, in contrast to that corresponding to the coding sequence of the $tRNA^{Arg}$ genes. Since the templates for sequencing were generated by PCR amplification, only the region between the primers, in this case the coding region of the $tRNA^{Arg}$ gene should be present to be sequenced. When the amplified DNA from aysmmetric PCR was analysed by agarose gel electrophoresis a number of faint bands were visible in addition to one strong band, which presumably corresponded to the specifically amplified single-stranded DNA product. It may be that the sequencing artifacts seen could have been avoided by purifying the major band from an agarose gel prior to sequencing. The conditions used for the asymmetric PCR (see section 2.5.15.2.) are prone to produce artifacts during amplification because of the large number of cycles involved (55 as opposed to 25-30 for a standard PCR amplification).

The cloning of two human $tRNA^{Gln}$ genes using the PCR was described in section 4.3.3. These genes were isolated as a result of a chance observation that primers specific for the $tRNA^{Gln}$ gene from λHt363 amplified two fragments of human genomic DNA rather than the one fragment expected. It may be that the use of tRNA gene-specific primers on genomic DNA could prove to be a rapid method of cloning tRNA genes where two isoacceptors are sufficiently close to each other in the genome, as in the case of the $tRNA^{Gln}$ genes mentioned above. The fact that some tRNA genes are clustered in the human genome might encourage such experiments. As Green *et al* ., (1990) pointed out, if two members of a tRNA gene family in close proximity in the genome lay in opposite orientations to each other then one primer would suffice to amplify the genomic fragment containing the genes, either the 5' or the 3' specific primer depending on the orientation of the genes. This method could be extended by using combinations of primers specific for different tRNA genes to attempt to amplify human genomic DNA fragments containing these genes. Recent work in our laboratory (A. Alloueche and J.P. Goddard, unpublished results) suggests that this approach might prove viable.

The use of the PCR as mentioned above is essentially a random method for cloning new tRNA genes, completely dependent on there being suitable combinations of genes close enough together to be amplified. A more systematic method for

potentially characterising all of the members of a human tRNA gene family would be to use inverse PCR. This technique can be used to amplify the regions flanking a known DNA sequence. The basic method (Ochman *et al* ., 1990) involves the digestion of target DNA with restriction enzymes and circularisation of the cleavage products to yield molecules that can act as templates for primers orientated in the opposite way to those normally employed for the PCR. Specifically, human genomic DNA would be digested with a restriction enzyme known to have no sites in a particular tRNA gene, and the intramolecularly ligated products subjected to an amplification reaction using tRNA gene-specific primers orientated so that extension proceeded outwards from the coding region of the gene. Restriction enzymes which cut human genomic DNA frequently would be the most suitable for this purpose since smaller target sequences allow more efficient amplification using the PCR. Different members of a tRNA gene family, in different genomic environments, would give rise to amplified fragments with sizes depending on the distance between the nearest flanking sites in the genome for the restriction enzyme used to cleave the genomic DNA. The net result would be a population of amplified fragments containing the immediate 5' and 3' flanking regions of all of the members of one tRNA gene family. The tRNA genes themselves could be isolated using the PCR once sequence data had been obtained from the flanking regions. The cloning of complete gene families would allow comparisons of flanking sequences and the possible correlation of such sequences with the *in vivo* and *in vitro* expression of the genes, in particular with tissue specific expression *in vivo* , perhaps by using the primer extension method described by Schmutzler and Gross (1990). The genomic distribution of the tRNA gene family could also be ascertained by the use of *in situ* hybridisation.

6.2. *In vitro* expression of the tRNA genes from λHt363.

6.2.1. *In vitro* transcription assays.

The *in vitro* transcription experiments detailed in Chapter five of this study gave internally consistent results which were in broad agreement with previous studies using this particular system (Gonos, 1989; Gonos and Goddard, 1990c). However,

the fact that the tRNA genes from λHt363 were transcriptionally active *in vitro* does not necessarily give an accurate reflection of their transcription *in vivo* and in particular does not address the question of tissue specific expression. The same is true for a heterologous *in vivo* transcription system such as *Xenopus* oocyte microinjection. It is possible to determine the levels of *in vivo* transcription of specific tRNA genes against the background of transcripts from other members of the same tRNA gene family. The primer extension method was used by Schmutzler and Gross, (1990) to study *in vivo* expression of human $tRNA^{Val}$ genes. However, this technique depends on different members of a gene family having a different immediate 5' flanking sequence in order to distinguish their primary transcripts. It would be interesting to see if the differences in transcriptional efficiency of the tRNA genes recorded in section 5.3. also occur *in vivo* , and if the same pattern exists in different tissues.

The limitations of an *in vitro* assay in the determination of transcription may also apply to the processing of transcripts in the system used, in that the results found in an artificial system may not give an accurate picture of what actually occurs in the whole organism. The precise nature of processing intermediates was not determined in this study so the processing pathway proposed for the $tRNA^{Arg}$ gene transcripts in section 5.5.1. was based purely on the changes in size of the processing intermediates. One method of specifically identifying intermediates would be to use hybridisation of oligonucleotides to tRNA precursor molecules isolated from a polyacrylamide gel. The synthesis of several different oligonucleotides specific for different parts of the precursor should allow rapid identification of processing intermediates, either through hybridisation or the primer extension method. Oligonucleotide primers have also been used to identify *in vivo* processing intermediates and splice junctions via PCR amplification (Delp *et al* ., 1991).

In the *in vitro* system used in this study, the four most tightly grouped tRNA genes cloned from the cluster showed the same relative transcription rates as separate genes as when transcribed together on a cloned restriction fragment from λHt363 (see section 5.4.). It might be expected that closely spaced tRNA genes in a cluster will compete for transcription factors *in vivo* . Since the internal promoters of tRNA genes are highly conserved it is probable that elements in the gene flanking sequences will be important if such competition occurs. When the separate tRNA genes from λHt363 were used in transcription assays they still had the same 5' and 3' flanking sequences as in the original cluster, spanning at least 100 residues upstream and at

least 60 downstream from the coding sequence. Because the immediate flanking sequences of the genes are the same even when the tRNA genes are taken out of the cluster, elements affecting transcriptional differences between the genes may still be present, explaining the maintenance of the relative transcriptional efficiencies.

It is not known whether there is any evolutionary or functional significance in the occurrence of tRNA gene clusters. It may be that clusters arose through duplication of an ancestral tRNA gene followed by sequence divergence. Alternatively, since retroposons have a tendancy to integrate into sites where other retroposons occur (Rogers, 1985; Weiner *et al* ., 1986) tRNA gene clusters may have arisen through several retroposition events involving different genes. It is also possible that tRNA gene clusters are selectively favoured because a group of genes together in the genome can sequester transcription factors more efficiently *in vivo* .

6.2.2. The role of the 5' flank of the tRNA^{Gly} gene from λHt363.

The experiments described in section 5.6. did not give a straightforward picture of the role of the 5' flanking sequence in transcription of the tRNA^{Gly} gene from λHt363. Several lines of evidence pointed towards part of this region being involved in controlling the transcription of the tRNA^{Gly} gene. The inversion of a 125 bp *Bam* HI fragment from the flank of the gene significantly reduces transcription, and this isolated fragment has been shown to compete, albeit not strongly, with other tRNA genes in a transcription assay. A larger region (175 bp, including the 125 bp *Bam* HI fragment) of the immediate 5' flank of the tRNA^{Gly} gene appears to bind to a protein or proteins in the HeLa cell extract used for transcription assays (see sections 5.6.3.). However, it is difficult to present a model which satisfactorily explains all of the results obtained. The 5' flank of the tRNA^{Gly} gene could be completely replaced with unrelated DNA (in recombinant pGLY5) without this having a large effect on transcription (see section 5.6.1.) and no footprint was observed when the 175 bp fragment of the tRNA^{Gly} gene 5' flank was used in DNaseI protection experiments (see section 5.6.3.). In addition, the isolated *Bam* HI fragment from the tRNA^{Gly} gene flank only competed significantly with tRNA genes in a transcription assay when used in a considerable molar excess (a 20-fold molar excess was needed to cause the same reduction in transcription of the tRNA^{Gly} gene in a competition assay

as that observed in a recombinant (pGLY3) with the 125 bp *Bam* HI fragment inverted).

The problem of relating the results of *in vitro* experiments to the actual situation *in vivo* has already been referred to several times in this chapter. Such problems may apply particularly to investigations into the role of flanking sequences in the control of tRNA gene expression since if regions outwith the gene sequence are involved in the modulation of transcription it is likely that in some cases the effect will be tissue-specific. Tissue-specific factors involved in the transcription of a particular tRNA gene might not be present in an *in vitro* transcription assay depending on the source of the cell-free extract used. However, the HeLa cell extract used in this study does appear to contain proteins that can associate with part of the $tRNA^{Gly}$ gene 5' flank.

In one set of experiments, described in section 5.6.4., 502 bp of the 5' flank of the $tRNA^{Gly}$ gene from λHt363 (from the *Pvu* II site to the residue immediately 5' of the coding sequence of the gene; see figure 5.9.) was placed in the 5' flank of the $tRNA^{Gln}$ gene from λHt363 in the recombinant pGLN2. This recombinant showed no difference in transcriptional efficiency from an unaltered $tRNA^{Gln}$ gene in the recombinant pGLN1 (see table 5.1.). However, pGLN2 retained 163 bp of the original immediate 5' flanking sequence of the $tRNA^{Gln}$ gene, so even if control elements are found in the $tRNA^{Gly}$ 5' flanking DNA they might have no effect when placed at this distance from a tRNA gene. More meaningful experiments would involve the complete replacement of the 5' flank of one tRNA gene with that of another. In the absence of convenient restriction sites such recombinants could be constructed by using the PCR to amplify the 5' flanking region of a tRNA gene, and cloning the resulting blunt ended fragments into the nearest suitable restriction site to the 5' end of another tRNA gene. Recombinants constructed in such a way could then be used in *in vitro* transcription assays. If such an approach were used for the tRNA genes in the cluster on λHt363 it should be possible to determine whether the observed differences in transcriptional efficiencies were caused by elements in the 5' flanks of the tRNA genes. The same approach could be used to look for control elements in the 3' flanking regions of the same genes.

REFERENCES

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

Adams, R.L.P. (1980) Cell Culture for Biochemists, Elsevier/N.Holland Biomedical Press.

Adams, R.L.P., Knowler, J.T. and Leader, D.P. (1986) The Biochemistry of the Nucleic acids, Chapman and Hall, London.

Adeniyi-Jones, S., Romeo, P.H. and Zasloff, M. (1984) *Nucleic Acids Res.* **12**, 1101-1115.

Allison, D.S., Goh, S.H. and Hall, B.D. (1983) *Cell* **34**, 655-664.

Allison, D.S. and Hall, B.D. (1985) *EMBO J.* **4**, 2657-2664.

Amstutz, H., Munz, P., Heyer, W.D., Leupold, U. and Kohli, J. (1985) *Cell* **40**, 879-886.

Arnold, G.J., Schmutzler, C., Thomann, U., Toll, H.V. and Gross, H.J. (1986) *Gene* **44**, 287-297.

Arnold, G.J. and Gross, H.J. (1987) *Gene* **51**, 237-246.

Arnold, G.J., Schmutzler, C. and Gross, H.J. (1988) *DNA* **7**, 87-97.

Atkinson, T. and Smith, M. (1984) in Oligonucleotides Synthesis, a Practical Approach, pp. 35-81, Gait,M.J.(ed.), I.R.L. Press.

Baker, R.E. and Hall, B.D. (1984) *EMBO J.* **3**, 2793-2800.

Baker, R.E., Gabrielsen, O. and Hall, B.D. (1986) *J. Biol. Chem.* **261**, 5275-5282.

Baker, R.E., Hall, B.D., Camier, S. and Sentenac, A. (1987) *Proc. Nat. Acad. Sci. U.S.A.* **84**, 8763-8767.

Baldi, M.I., Mattocia, E., Ciafre, S., Attardi, D.G. and Tocchini-Valentini, G.P. (1986) *Cell* **47**, 965-971.

Bartholomew, B., Kassavetis, G.A., Braun, B.R. and Geiduschek, E.P. (1990) *EMBO J.* **9**, 2197-2205.

Besser, D., Gotz, F., Schulze-Forster, K., Wagner, H., Kroger, H. and Simon, D. (1990) *FEBS Lett.* **269**, 358-362.

Bieker, J.J., Martin, P.L. and Roeder, R.G. (1985) *Cell* **40**, 119-127.

Bird, A.P. (1978) *J. Mol. Biol.* **118**, 49-60.

Bird, A.P. (1986) *Nature* **321**, 209-213.

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

Bogenhagen, D.F., Sakonju, S. and Brown, D.D. (1980) *Cell* **19**, 27-35.

Bogenhagen, D.F., Wormington, W.M. and Brown, D.D. (1982) *Cell* **28**, 413-421.

Bogenhagen, D.F. (1985) *J. Biol. Chem.* **260**, 6466-6471.

Bothwell, A. (1990) in Methods for Cloning and Analysis of Eukaryotic Genes, pp. 177-179, Bothwell, A., Yancopoulos, G.D. and Alt, F.W. (eds.), Jones and Bartlett, Boston.

Boyd, E., Theriault, A., Goddard, J.P., Kalaitsidaki, M., Spathas, D.H. and Connor, J.M., (1989) *Hum. Genet.* **81**, 153-156.

Bradford, M.M. (1976) *Anal. Biochem.* **72**, 248-254.

Brodeur, G.M., Sandameyer, S.B. and Olson, M.W. (1983) *Proc. Nat. Acad. Sci. U.S.A.* **80**, 3292-3296.

Brown, D.D. (1984) *Cell* **37**, 359-365.

Buckland, R.A., Cooke, H.J., Roy, K.L., Dahlberg, J.E. and Lund, E. (1983) *Gene* **22**, 211-217.

Buckland, R.A. (1989) *Am. J. Hum. Genet.* **45**, 283-295.

Burke, D.J., Schaak, J., Sharp, S., Soll, D., (1980) *J. Biol. Chem.* **258**, 15224-15231.

Burke, D.J. and Soll, D. (1985) *J. Biol. Chem.* **260**, 816-823.

Camier, S., Gabrielsen, O., Baker, R.E. and Sentenac, A. (1985) *EMBO J.* **4**, 491-500.

Camier, S. Baker, R.E. and Sentenac, A. (1990) *Nucleic Acids Res.* **18**, 4571-4578.

Carey, M.F., Gerrard, S.P. and Cozzarelli, N.R. (1986) *J. Biol. Chem.* **261**, 4309-4317.

Carrara, G., Di Segni, G., Otsuka, A. and Tocchini-Valentini, G.P. (1981) *Cell* **27**, 371-379.

Castagnoli, L., Ciliberto, G. and Cortese, R. (1982) *Nucleic Acids Res.* **10**, 4135-4145.

Chang, Y.N., Pirtle, I.L. and Pirtle, R.M. (1986) *Gene* **48**, 165-174.

Choffat, Y., Suter, B. Behra, R. and Kubli, E. (1988) *Mol. Cell. Biol.* **8**, 3332-3337.

Church, G.M. and Gilbert, W. (1984) *Proc. Nat. Acad. Sci. U.S.A.* **81**, 1991-1995.

Ciliberto, G., Traboni, C. and Cortese, R. (1982) *Proc. Nat. Acad. Sci. U.S.A.* **79**, 1921-1925.

Ciliberto, G., Castagnoli, L. and Cortese, R. (1983a) *Curr. Topics Dev. Biol.* **18**, 59-87.

Ciliberto, G., Costanzo, R.F., Dente, L. and Cortese, R. (1983b) *Cell* **32**, 725-733.

Clark, B.F.C. and Klug, A. (1975) *Proc. 10th FEBS meeting, Paris*, 183.

Clarkson, S.G. and Kurer, V. (1976) *Cell* **8**, 183-195.

Clarkson, S.G., Kurer, V. and Smith, H.O. (1978) *Cell* **14**, 713-724.

Clarkson, S.G. (1983) in Eukaryotic genes: their structure, activity and regulation, pp 239-261, Maclean, N., Gregory, S.P. and Flavell R.A. (eds.), Butterworths Press, London.

Cooley, L., Schaack, J., Burke, D.J., Thomas, B. and Soll, D. (1984) *Mol. Cell. Biol.* **4**, 2714-2722.

Craig, L.C., Wang, L.P., Lee, M.M., Pirtle, I.L. and Pirtle, R.M. (1989) *DNA* **8**, 457-471.

Cribbs, D.L., Leung, J., Newton, C.H., Hayashi, S., Miller, R.C. and Tener, G.M. (1987) *J. Mol. Biol.* **197**, 397-404.

Dagert, M. and Erlich, S.D. (1979) *Gene* **6**, 23-28.

De Franco, R., Schmidt, O. and Soll, D. (1980) *Proc. Nat. Acad. Sci. U.S.A.* **77**, 3365-3368.

De Franco, D., Sharp, S. and Soll, D. (1981) *J. Biol. Chem.* **256**, 12424-12429.

De Robertis, E.M. and Olson, M.V. (1979) *Nature* **278**, 137-143.

De Robertis, E.M., Black, P. and Nishikura, K. (1981) *Cell* **23**, 89-93.

Dean, N. and Berk, A.J. (1987) *Nucleic Acids Res.* **15**, 9895-9907.

Dean, N. and Berk, A.J. (1988) *Mol. Cell. Biol.* **8**, 3017-3025.

Del Rey, F.J., Donahue, T.F. and Fink, G.R. (1982) *Proc. Nat. Acad. Sci. U.S.A.* **79**, 4138-4142.

Delp, G., Igloi, G.L. and Kossel, H. (1991) *Nucleic Acids Res.* **19**, 713-716.

Deutscher, M.P. (1984) *Crit. Rev. Biochem.* **17**, 45-71.

Dingermann, T., Burke, D.J., Sharp, S., Schaack, J. and Soll, D. (1982) *J. Biol. Chem.* **257**, 14738-14744.

Dingermann, T., Sharp, S., Schaack, J. and Soll, D. (1983) *J. Biol. Chem.* **258**, 10395-10402.

Dingermann, T. and Nerke, K. (1987) *Anal. Biochem.* **162**, 466-475.

Doering, J.L., Jelachich, M.L. and Hanlin, K.M. (1982) *FEBS Lett.* **146**, 47-51.

Doran, J.L., Wei, X. and Roy, K.L. (1987) *Gene* **56**, 231-243.

Doran, J.L., Bingle, W.H. and Roy, K.L. (1988) *Gene* **65**, 329-336.

Drabkin, H.J. and RajBhandary, U.L. (1985a) *J. Biol. Chem.* **260**, 5588-5595.

Drabkin, H.J. and RajBhandary, U.L. (1985b) *J. Biol. Chem.* **260**, 5596-5602.

Eigel, A. and Feldmann, H. (1982) *EMBO J.* **10**, 1245-1250.

Engelke, D.R., Ng, S.Y., Shastry, B.S. and Roeder, R.G. (1980) *Cell* **19**, 717-728.

Fabrizio, P., Coppo, A., Fruscoloni, P., Benedetti, P., Di Segni, G. and Tocchini-Valentini, G.P. (1987) *Proc. Nat. Acad. Sci. U.S.A.* **84**, 8763-8767.

Feinberg, A.R. and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6-13.

Feinberg, A.R. and Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266-267.

Folk, W.R. and Hofstetter, H. (1983) *Cell* **33**, 585-593.

Folk, W.R. (1988) *Genes and Development* **2**, 373-375.

Fostel, J., Narayanswani, S., Hankalo, B., Clarkson, S.G. and Pardlk, M.L. (1984) *Chromosoma* **90**, 254-260.

Fuhrman, S.A., Engelke, D.R. and Geiduschek, E.P. (1984) *J. Biol. Chem.* **259**, 1934-1943.

Gabrielsen, O.S. and Oyen, T.B. (1987) *Nucleic Acids Res.* **15**, 5699-5713.

Gabrielsen, O.S., Hornes, E., Korsnes, L., Ruet, A. and Oyen, T.B. (1989a) *Nucleic Acids Res.* **17**, 6253-6267.

Gabrielsen, O.S., Marzouki, N., Ruet, A., Sentenac, A. and Fromageot, P. (1989b) *J. Biol. Chem.* **264**, 7505-7511.

Gall, J.G. and Pardue, M.L. (1969) *Proc. Nat. Acad. Sci. U.S.A.* **63**, 378-382.

Galli, G., Hofstetter, H. and Birnstiel, M.L. (1981) *Nature* **294**, 626-631.

Gamulin, V., Mao, J., Appel, B., Smith, M.S., Yamao, F. and Soll, D. (1983)

Nucleic Acids Res. **11**, 8537-8546.

Garber, R.L. and Gage, L.P. (1979) *Cell* **18**, 817-828.

Geiduschek, E.P. and Tocchini-Valentini, G.P. (1988) *Annu. Rev. Biochem.* **57**, 873-914.

Gibson, T.J. (1984) Ph.D. Thesis, Cambridge University, England.

Glew, L., Lo, R., Reece, T., Nichols, M., Soll, D. and Bell, J. (1986) *Gene* **44**, 307-314.

Goddard, J.P., Squire, M., Bienz, M. and Smith, J.D. (1983) *Nucleic Acids Res.* **11**, 2551-2561.

Gonos, E.S. (1989) Ph.D. Thesis, Glasgow University, Scotland.

Gonos, E.S. and Goddard, J.P. (1990a) *FEBS Lett.* **276**, 138-142.

Gonos, E.S. and Goddard, J.P. (1990b) *Nucleic Acids Res.* **18**, 6705.

Gonos, E.S. and Goddard, J.P. (1990c) *Biochem. J.* **272**, 797-803.

Gottesfeld, J. and Bloomer, L.S. (1982) *Cell* **28**, 781-791.

Gouilloud, E. and Clarkson, S.G. (1986) *J. Biol. Chem.* **261**, 486-494.

Green, C.J., Sohel, I. and Vold, B.S. (1990) *J. Biol. Chem.* **265**, 12139-12142.

Guilfoyle, R. and Weinmann, R. (1981) *Proc. Nat. Acad. Sci. U.S.A.* **78**, 3378-3382.

Gupta, C., Roe, B.A. and Randerath, K. (1979) *Nucleic Acids Res.* **7**, 959-970.

Hall, B.J., Clarkson, S.G. and Tocchini-Valentini, G. (1982) *Cell* **29**, 3-5.

Han, J.H. and Harding, J.D. (1982) *Nucleic Acids Res.* **10**, 4891-4890.

Han, J.H., Rooney, R.J. and Harding, J.D. (1984) *Gene* **28**, 249-255.

Hatfield, D., Lee, B.J., Hampton, L. and Diamond, A.M. (1991) *Nucleic Acids Res.* **19**, 939-943.

Hatlen, L. and Attardi, G. (1971) *J. Mol. Biol.* **56**, 535-553.

Hauber, J., Stucka, R., Krieg, R. and Feldmann, H. (1988) *Nucleic Acids Res.* **16**, 10623-10634.

Hegg, L.A. and Thurlow, D.L. (1990) *Nucleic Acids Res.* **18**, 2993-3000.

Hershey, N.D. and Davidson, N. (1980) *Nucleic Acids Res.* **8**, 4899-4910.

Himeno, H., Tsunemi, H., Ueda, T., Watanabe, K. and Shimizu, M. (1990) *Nucleic Acids Res.* **18**, 6815-6819.

Hipskind, R.A. and Clarkson, S.G. (1983) *Cell* **34**, 881-890.

Hoagland, M.B., Stephenson, M.L., Scott, J.F., Hecht, L.I. and Zamecnik, P.C.

(1958) *J. Biol. Chem.* **231**, 241-257.

Hoe, K.L., Hong, H.J., Yoo, S.H. and Yoo, O.J. (1987) *Nucleic Acids Res.* **15**, 10045.

Hofstetter, H., Kressmann, A. and Birnstiel, M.L. (1981) *Cell* **24**, 573-585.

Hohn, B. (1979) *Methods Enzymol.* **68**, 299-309.

Holley, R.W., Apgar, J., Evesett, G.A., Madison, J.T., Marquisee, M., Marill, S.H., Penswick, J.R. and Zamir, A. (1965) *Science* **147**, 1462-1465.

Holmes, D.S. and Quigley, M. (1981) *Anal. Biochem.* **114**, 193-197.

Hong, H.J., Yoo, S.H. and Yoo, O.J. (1987) *Nucleic Acids Res.* **15**, 4987.

Horvath, D. and Spiegelman, G.B. (1988) *Nucleic Acids Res.* **16**, 2585-2599.

Hosbach, H.A., Silberklang, M. and Mc Carthy, B.J. (1980) *Cell* **21**, 169-178.

Hou, Y.M. and Schimmel, P. (1988) *Nature* **333**, 140-145.

Hovemann, B., Sharp, S., Yamada, H. and Soll, D. (1980) *Cell* **19**, 889-895.

Hu, J.C., Lund, E. and Dahlberg, J.E. (1983) *Nucleic Acids Res.* **11**, 4809-4821.

Huibregtse, J.M., Evans, C.F. and Engelke, D.R. (1987) *Mol. Cell. Biol.* **7**, 3212-3220.

Innes, M.A. and Gelfand, D.H. (1990) in PCR Protocols, A Guide to Methods and Applications, pp. 3-12, Innes, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (eds), Academic Press, California.

Jacob, S.T. (1973) *Prog. Nucleic Acids Res. Mol. Biol.* **13**, 93-126.

Johnson, D.L. and Wilson, S.L. (1989) *Mol. Cell. Biol.* **9**, 2018-2024.

Johnson, J.D. and Raymond, G.J. (1984) *J. Biol. Chem.* **259**, 5990-5994.

Johnson, J.D., Raymond, G.J. and De Parasis, J. (1984) *Mol. Gen. Genet.* **197**, 55-61.

Johnson, P.F. and Abelson, J. (1983) *Nature* **302**, 681-687.

Kariya, Y., Kato, K., Hayashizaki, Y., Hineno, S., Tarui, S. and Matsubara, K. (1987) *Gene* **53**, 1-10.

Kassavetis, G.A., Riggs, D.L., Negri, R., Nguyen, L.H. and Geiduschek, E.P. (1989) *Mol. Cell. Biol.* **9**, 2551-2566.

Kassavetis, G.A., Braun, B.R., Nguyen, L.H. and Geiduschek, E.P. (1990) *Cell* **60**, 235-245.

Kim, S.H., Sussman, J.C., Suddath, F.L., Quigley, G.J., McPherson, A., Wang, A.H.J., Seeman, N.C. and Rich, A. (1974) *Proc. Nat. Acad. Sci. U.S.A.*

71, 4970-4974.

Kim, S.H. (1979) in Transfer RNA : Structure, Properties and Recognition, pp.83-100, Schimmel, P.R., Soll, D. and Abelson J.N. (eds.) Cold Spring Harbour Laboratory, CHS New York.

Klekamp, M.S. and Weil, P.A. (1986a) *J. Biol. Chem.* **261**, 2819-2827.

Klekamp, M.S. and Weil, P.A. (1986b) *Arch. Biochem. Biophys.* **246**, 783-800.

Klekamp, M.S. and Weil, P.A. (1987) *J. Biol. Chem.* **262**, 7878-7883.

Klemenz, R., Stillman, D.J. and Geiduschek, E.P. (1982) *Proc. Nat. Acad. Sci. U.S.A.* **79**, 6191-6195.

Knapp, G., Beckmann, J.S., Johnson, P.F., Fuhrmann, S.A. and Abelson, J. (1978) *Cell* **14**, 221-236.

Koski, R.A., Clarkson, S.G., Kurjan, J., Hall, B.D. and Smith, M. (1980) *Cell* **22**, 415-425.

Koski, R.A. and Clarkson, S.G. (1982) *J. Biol. Chem.* **257**, 4514-4521.

Koski, R.A., Allison, D.S., Worthington, M. and Hall, B.D. (1982) *Nucleic Acids Res.* **10**, 8127-8143.

Kressmann, A., Hofstetter, H., Di Capua, E., Grosschedl, R. and Birnstiel, M.L. (1979) *Nucleic Acids Res.* **7**, 1749-1763.

Krupp, J.L., Shu, H.H. and Martin, N.C. (1988) *Nucleic Acids Res.* **16**, 770.

Larson, D., Bradford-Wilcox, J., Young, L.S. and Sprague, K.U. (1983) *Proc. Nat. Acad. Sci. U.S.A.* **80**, 3416-3420.

Laski, F.A., Fire, A.Z., RajBhandary, U.L. and Sharp, P.A. (1983) *J. Biol. Chem.* **258**, 11974-11980.

Lassar, A.B., Martin, P.L. and Roeder, R.G. (1983) *Science* **222**, 740-748.

Lasser-Weis, M., Bawnick, N., Rosen, A., Sarid, S. and Daniel, V. (1981) *Nucleic Acids Res.* **9**, 5965-5978.

Lawn, R.M., Fritch, E.F., Parker, R.C., Blake, G. and Maniatis, T. (1978) *Cell* **15**, 1157-1174.

Lin, V.K. and Agris, P.F. (1980) *Nucleic Acids Res.* **8**, 3467-3480.

Lofquist, A. and Sharp, S. (1986) *J. Biol. Chem.* **261**, 14600-14606.

Lofquist, A., Garcia, A. and Sharp, S.J. (1988) *Mol. Cell. Biol.* **8**, 4441-4449.

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

Looney, J.E. and Harding, J.D. (1983) *Nucleic Acids Res.* **11**, 8761-8775.

Ma, D.P., Lund, E., Dahlberg, J.E. and Roe, B.A. (1984) *Gene* **28**, 257-262.

Makowski, D.R., Haas, R.A., Dolan, K.P. and Grunberger, D. (1983) *Nucleic Acids Res.* **11**, 8609-8624.

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978) *Cell* **15**, 687-701.

Mao, J., Schmidt, O. and Soll, D. (1980) *Cell* **21**, 509-516.

Marchfelder, A., Schuster, W. and Brennicke, A. (1990) *Nucleic Acids Res.* **18**, 1401-1406.

Marschalek, R. and Dingermann, T. (1988) *Nucleic Acids Res.* **16**, 6737-6752.

Martin, R. (1987) *BRL Focus* **9**, 7-8.

Marzluff, W.F., White, E.L., Benjamin, R. and Huang, R.C.C. (1975) *Biochemistry* **14**, 3715-3724.

Marzouki, N., Camier, S., Ruet, A., Moenne, A. and Sentenac, A. (1986) *Nature* **323**, 176-178.

Mattocia, E., Baldi, M.I., Pande, G., Ogden, R. and Tocchini-Valentini, G.P. (1983) *Cell* **32**, 67-76.

McBride, O.W., Rajagopalan, M. and Hatfield, D. (1987) *J. Biol. Chem.* **262**, 11163-11166.

McBride, O.W., Pirtle, I.L. and Pirtle, R.M. (1989) *Genomics* **5**, 561-573.

McCabe, P.C. (1990) in PCR Protocols, A Guide to Methods and Applications, pp. 76-83, Innes, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (eds), Academic Press, California.

McLaren, A. and Goddard, J.P. (1986) *Nucleic Acids Res.* **14**, 5938.

McPherson, J.M. and Roy, K.L. (1986) *Gene* **42**, 101-106.

Melton, D.A. and Cortese, R. (1979) *Cell* **18**, 1165-1172.

Melton, D.A., De Robertis, E.M. and Cortese, R. (1980) *Nature* **284**, 143-148.

Miller, S.A., Dykes, D.D. and Polesky, H.F. (1988) *Nucleic Acids Res.* **16**, 1215.

Morrison, N., Goddard, J.P., Boyd, E., Bourn, D. and Conner, J.M. (1991) *Hum. Genet.* (in press).

Morry, M.J. and Harding, J.D. (1986) *Mol. Cell. Biol.* **6**, 105-115.

Morton, D.G. and Sprague, K.U. (1984) *Proc. Nat. Acad. Sci. U.S.A.* **81**, 5519-5522.

Muller, F. and Clarkson, S.G. (1980) *Cell* **19**, 345-353.

Muller, F., Clarkson, S.G. and Galas, D.J. (1987) *Nucleic Acids Res.* **15**, 7191.

Murphy, M.H. and Baralle, F.E. (1983) *Nucleic Acids Res.* **11**, 7695-7700.

Naylor, S.L., Sakagushi, A.Y., Shows, T.B., Grzeschik, K.H., Holmes, M. and Zasloff, M. (1983) *Proc. Nat. Acad. Sci. U.S.A.* **80**, 5027-5031.

Newman, A.J., Ogden, R.C. and Abelson, J. (1983) *Cell* **35**, 117-125.

Nordheim, A. and Rich, A. (1983) *Nature* **303**, 674-679.

Normanly, J., Ogden, R.C., Horvath, S.J. and Abelson, J. (1986) *Nature* **321**, 213-219.

Normanly, J. and Abelson, J. (1989) *Annu. Rev. Biochem.* **58**, 1029-1049.

Norrander, J., Kempe, T. and Messing, J. (1983) *Gene* **26**, 101-106.

O'Connor, J.P. and Peebles, C.L. (1991) *Mol. Cell. Biol.* **11**, 425-439.

O'Farrel, P.Z., Cordell, B., Valenzuela, P., Rutter, W.J. and Goodman, H.M. (1978) *Nature* **274**, 438-445.

O'Neil, V.A., Eden, F.C., Pratt, K. and Hatfield, D.L. (1985) *J. Biol. Chem.* **260**, 2501-2508.

Ochman, H., Medhora, M.M., Garza, D. and Hartl, D.L. (1990) in PCR Protocols, A Guide to Methods and Applications, pp. 219-227, Innes, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (eds), Academic Press, California.

Ogden, R.C., Lee, M.C. and Knapp, G. (1984) *Nucleic Acids Res.* **12**, 9367-9382.

Olsen, S., Pirtle, I.L., Pirtle, R.M. and McBride, O.W. (1987) *Am. J. Hum. Genet.* **41**, (supplement) A180.

Palmer, J.M. and Folk, W.R. (1990) *T.I.B.S.* **15**, 300-304.

Peterson, R.C. (1987) *Biochem. Biophys. Acta.* **908**, 81-89.

Pieler, T., Appel, B., Oei, S.L., Mentzel, H. and Erdmann, V.A. (1985a) *EMBO J.* **4**, 1847-1853.

Pieler, T., Oei, S.L., Hamm, J., Engelke, U. and Erdmann, V.A. (1985b) *EMBO J.* **4**, 3751-3756.

Pieler, T., Hamm, J. and Roeder, R.G., (1987) *Cell* **48**, 91-100.

Pirtle, I.L., Shortridge, R.D. and Pirtle, R.M. (1986) *Gene* **43**, 155-167.

Raymond, G.J. and Johnson, J.D. (1983) *Nucleic Acids Res.* **11**, 5969-5988.

Raymond, G.J. and Johnson, J.D. (1987) *Nucleic Acids Res.* **15**, 9881-9894.

Raymond, K.C., Raymond, G.J. and Johnson, J.D. (1985) *EMBO J.* **4**,

2649-2656.

Reilly, J.G., Ogden, R. and Rossi, J.J. (1982) *Nature* **300**, 287-289.

Rich, A. and RajBhandary, U.L. (1976) *Annu. Rev. Biochem.* **45**, 805-860.

Robertus, J.D., Ladner, J.E., Finch, J.T., Rhodes, D., Brown, R.S., Clark, B.F.C. and Klug A. (1974) *Nature* **250**, 546-551.

Rogers, J.H. (1985) *Int. Rev. Cytol.* **93**, 187-279.

Rooney, R.J. and Harding, J.D. (1986) *Nucleic Acids Res.* **14**, 4849-4864.

Rooney, R.J. and Harding, J.D. (1988) *Nucleic Acids Res.* **16**, 2509-2521.

Rosen, A., Sarid, S. and Daniel, V. (1984) *Nucleic Acids Res.* **12**, 4893-4906.

Rosen, A. and Daniel, V. (1988) *Gene* **69**, 275-285.

Roy, K.L., Cooke, H. and Buckland, R. (1982) *Nucleic Acids Res.* **10**, 7313-7322.

Ruet, A., Camier, S., Smagowiez, W., Sentenac, A. and Fromageot, P. (1984) *EMBO J.* **3**, 343-350.

Russo, T., Duilio, A., Ammendola, R., Costanzo, F. and Cimino, F. (1987a) *Nucleic Acids Res.* **15**, 8562.

Russo, T., Oliva, A., Duilio, A., Ammendola, R., Costanzo, F., Zannini, M. and Cimino, F. (1987b) *Bioch. Bioph. Res. Commun.* **149**, 1118-1124.

Sajjadi, F.G. and Spiegelman, G.B. (1987) *Gene* **60**,13-19.

Sajjadi, F.G., Miller, R.C. and Spiegelman, G.B. (1987) *Mol. Gen. Genet.*. **206**, 279-284.

Sakonju, S., Bogenhagen, D.F. and Brown, D.D. (1980) *Cell* **19**, 13-25.

Sakonju, S., Brown, D.D., Engelke, D.R., Ng, S.Y., Shastry, B.S. and Roeder, R.G. (1981) *Cell* **23**, 665-669.

Saluz, H. and Jost, J.P. (1989) *Proc. Nat. Acad. Sci. U.S.A.* **86**, 2602-2606.

Sambrook, J. Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, C.S.H. New York.

Sampson, J.R., Di Renzo, A., Behlen, L. and Uhlenbeck, O.C. (1989) *Science* **243**, 1363-1366.

Sandmeyer, S.B. and Olson, M.V. (1982) *Proc. Nat. Acad. Sci. U.S.A.* **79**, 7674-7678.

Sandmeyer, S.B., Bilanchone, V.W., Clark, D.J., Marcos, P., Carle, G.F. and Brodeur, G.M. (1988) *Nucleic Acids Res.* **16**, 1499-1515.

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

Santoro, C., Costanzo, F. and Ciliberto, G. (1984) *EMBO J.* **3**, 1553-1559.

Santos, T. and Zasloff, M. (1981) *Cell* **23**, 699-709.

Schaack, J., Sharp, S., Dingermann, T. and Soll, D. (1983) *J. Biol. Chem.* **258**, 2447-2453.

Schaack, J., Sharp, S., Dingermann, T., Burke, D.J., Cooley, L. and Soll, D. (1984) *J. Biol. Chem.* **259**, 1461-1467.

Schaack, J. and Soll, D. (1985) *Nucleic Acids Res.* **13**, 2803-2814.

Schlissel, M.S. and Brown, D.D. (1984) *Cell* **37**, 903-913.

Schmidt, O., Mao, J., Ogden, R., Beckmann, J., Sakano, H., Abelson, J. and Soll, D. (1980) *Nature* **287**, 750-752.

Schmidtke, J. and Cooper, D.N. (1990) *Nucleic Acids Res.* **18**, (sequences supplement) 2413-2547.

Schmutzler, C. and Gross, H.J. (1990) *Nucleic Acids Res.* **18**, 5001-5008.

Schneider, H.R., Waldschmidt, R., Jahn, D. and Seifart, K.H. (1989) *Nucleic Acids Res.* **17**, 5003-5017.

Schneider, H.R., Waldschmidt, R. and Seifart, K.H. (1990) *Nucleic Acids Res.* **18**, 4743-4750.

Schulman, L.H. and Pelka, H. (1988) *Science* **242**, 765-768.

Schulman, L.H. and Pelka, H. (1990) *Nucleic Acids Res.* **18**, 285-289.

Schweizer, E., MacKennie, C. and Halvorsen, H.O. (1969) *J. Mol. Biol.* **40**, 261-277.

Segall, J., Matsui, T. and Roeder, R.G. (1980) *J. Biol. Chem.* **255**, 11986-11991.

Sekiya, T., Kuchino, Y. and Nishimura, S. (1981) *Nucleic Acids Res.* **9**, 2239-2250.

Setzer, D.R. and Brown, D.D. (1985) *J. Biol. Chem.* **260**, 2483-2492.

Sharp, S., De Franco, D., Silberklang, M., Hosbach, H.A., Schmidt, T., Kubli, E., Gergen, J.P., Wensink, P.C. and Soll, D. (1981a) *Nucleic Acids Res.* **9**, 5867-5882.

Sharp, S., De Franco, D., Dingermann, T., Farell, P. and Soll, D. (1981b) *Proc. Nat. Acad. Sci. U.S.A.* **78**, 6657-6661.

Sharp, S., Dingermann, T. and Soll, D. (1982) *Nucleic Acids Res.* **10**, 5393-5406.

Sharp, S., Dingermann, T., Schaack, J., De Franko, D. and Soll, D. (1983a) *J.Biol. Chem.* **258**, 2440-2446.

Sharp, S., Dingermann, T., Schack, J., Sharp, J.A., Burke, D.J., De Robertis, E.M. and Soll, D. (1983b) *Nucleic Acids Res.* **11**, 8677-8690.

Sharp, S., Schaack, J., Cooley, L., Burke, D.J. and Soll, D. (1985) *Crit. Rev. Biochem.* **19**, 107-144.

Shaw, K.J. and Olson, K.V. (1984) *Mol. Cell. Biol.* **4**, 657-665.

Shibuya, K., Noguchi, S., Nishimura, S. and Sekiya, T. (1982) *Nucleic Acids Res.* **10**, 4441-4448.

Shibuya, K., Nogushi, S., Yamaki, M., Nishimura, S. and Sekiya, T. (1985) *J.Biochem.* **97**, 1719-1725.

Shortridge, R.D., Pirtle, I.L. and Pirtle, R.M. (1985) *Gene* **33**, 269-277.

Shortridge, R.D., Johnson, G.D., Craig, L.C., Pirtle, I.M. and Pirtle, R.M. (1989) *Gene* **79**, 309-324.

Smardo, F.L. and Calvet, J.P. (1987) *Gene* **57**, 213-220.

Sollner-Webb, B. (1988) *Cell* **52**, 153-154.

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

Sprague, K.U., Larson, D., and Morton, D. (1980) *Cell* **22**, 171-178.

Sprinzl, M., Hartmann, T., Weber, J., Blank, J. and Zeidler, R. (1989) *Nucleic Acids Res.* **17**, (sequences supplement) r1-r172.

Staden, R. (1986) *Nucleic Acids Res.* **14**, 217-231.

Standring, D.N., Venegas, A. and Rutter, W.J. (1981) *Proc. Nat. Acad. Sci. U.S.A.* **78**, 5963-5967.

Steffenson, D.M. and Wimber, D.E. (1971) *Genetics* **69**, 163-178.

Stillman, D.J. and Geiduschek, E.P. (1984) *EMBO J.* **3**, 847-853.

Stillman, D.J., Sivertsen, A.L., Zetner, P.G. and Geiduschek, E.P. (1984) *J. Biol. Chem.* **259**, 7955-7962.

Stillman, D.J., Better, M. and Geiduschek, E.P. (1985) *J. Mol. Biol.* **185**, 451-455.

Stuka, R. and Feldmann, H. (1988) *Nucleic Acids Res.* **16**, 3583.

Suddath, F.L., Quiglely, J., McPherson, A., Sneden, D., Kim, J., Kim, S.H. and Rich, A. (1974) *Nature* **248**, 20-24.

Suter, B. and Kubli, E. (1988) *Mol. Cel. Biol.* **8**, 3322-3331.

Swerdlow, H. and Guthrie, C. (1984) *J. Biol. Chem.* **259**, 5197-5207.

Szekely, E., Belford, H.G. and Greer, C.L. (1988) *J. Biol. Chem.* **263**, 13839-13847.

Szweykowska-Kulinska, Z. and Beier, H. (1991) *Nucleic Acids Res.* **19**, 707-712.

Thomann, H., Schmutzler, C., Hudepohl, U., Blow, M. and Gross, H.J. (1989) *J. Mol. Biol.* **209**, 505-523.

Thurlow, D.L., Shilowski, D. and Marsh, T.L. (1991) *Nucleic Acids Res.* **19**, 885-891.

Traboni, C., Cortese, R. and Salvatore, F. (1980) *Nucleic Acids Res.* **8**, 5223-5232.

Traboni, C., Ciliberto, G. and Cortese, R. (1984) *Cell* **36**, 179-187.

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

van Dyke, M.W. and Roeder, R.G. (1987) *Nucleic Acids Res.* **15**, 5031-5039.

van Tol, H., Stange, N., Gross, H.J. and Beier, H. (1987) *EMBO J.* **6**, 35-42.

van Tol, H. and Beier, H. (1988) *Nucleic Acids Res.* **16**, 1951-1966.

Vnencak-Jones, C.L., Wahab, S.Z., Zehener, Z.E. and Holmes, W.M. (1987) *Mol. Cell. Biol.* **7**, 4134-4138.

Wahab, S.Z., Holmes, W.M. and Zehner, Z.E. (1989) *Gene* **77**, 361-370.

Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A. and Weiner, A.M. (1987) Molecular Biology of the Gene, The Benjamin Publishing Company.

Weber, L. and Berger, E. (1976) *Biochemistry* **15**, 5511-5519.

Weil, P.A., Segall, J., Harris, B., Ng, S.Y. and Roeder, R.G. (1979) *J. Biol. Chem.* **254**, 6163-6173.

Weiner, A.M., Deninger, P.L. and Efstratiadis, A. (1986) *Annu. Rev. Biochem.* **55**, 631-661.

White, T.J., Arnheim, N. and Erlich, H.A. (1989) *T.I.G.* **5**, 185-189.

Willis, J., Hottinger, H., Pearson, D., Chisholm, V., Leupold, U. and Soll, D. (1984) *EMBO J.* **3**, 1573-1580.

Wilson, E.T., Larson, D., Young, L.S. and Sprague, K.U. (1985) *J. Mol. Biol.* **183**, 153-163.

Wormington, W.M., Bogenhagen, D.F., Jordan, E. and Brown, D.D. (1981) *Cell* **24**, 809-817.

Wu, G.J., Railley, J.F. and Cannon, R.E. (1987) *J. Mol. Biol.* **194**, 423-442.

Yanish-Perron, C., Viera, J. and Messing, J. (1985) *Gene* **33**, 103-119.

Yen, P.H. and Davidson, N. (1980) *Cell* **22**, 137-148.

Yoo, S.H. (1984) *Korean J. Genet.* **6**, 101-108.

Yoshinaga, S.K., Boulanger, P.A. and Berk, A.J. (1987) *Proc. Nat. Acad. Sci. U.S.A.* **84**, 3585-3589.

Yoshinaga, S.K., L'Etoile, N.D. and Berk, A.J. (1989) *J. Biol. Chem.* **264**, 10726-10731.

Young, L.S., Takahashi, N. and Sprague, K. U. (1986) *Proc. Nat. Acad. Sci. U.S.A.* **83**, 374-378.

Zasloff, M. and Santos, T. (1980) *Proc. Nat. Acad. Sci. U.S.A.* **77**, 5668-5672.

Zasloff, M., Santos, T., Romeo, P. and Rosenberg, M. (1982) *J. Biol. Chem.* **257**, 7857-7863.

Publications

Bourn, D. and Goddard, J.P. (1990) *Nucleic Acids Res.* **18**, 7442.

Morrison, N., Goddard, J.P., Boyd, E., Bourn, D. and Connor, J.M. (1991) *Hum. Genet.* (in press).