SEQUENCE ANALYSIS OF A tRNA GENE CLUSTER:

DROSOPHILA LEUCINE-tRNA GENES CONTAIN INTERVENING SEQUENCES

Thesis by

Randy Richard Robinson

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1981

(submitted August 22, 1980)

To my parents, for motivation.

To my advisor, for inspiration.

To Debby and Erin, for understanding.

Acknowledgements

I owe a great deal to my advisor, Norman Davidson, in whose laboratory this work was accomplished. I shall not forget his dedication to his work and his colleagues, or his intense desire to discover and test new ideas. The many postdoctoral fellows and graduate students in Dr. Davidson's laboratory provided a stimulating environment of scientific discussion and exchanges. I especially thank Dr. Pauline Yen for the original gift of Ch4:Dm50AB and instruction in Southern hybridization and *in situ* hybridization, Dr. N. Davis Hershey for introducing me to Maxam-Gilbert sequencing, and Dr. Kenshi Hayashi for helpful discussions on the sequencing of tRNA genes. I also thank the members of the Tom Maniatis group for helpful discussions, especially Dr. Ed Fritsch, who gave me valuable instruction and extracts for *in vitro* transcription, and Dr. Pamela Mellon, for instruction in DEAE paper electrophoresis. Financial aid was provided by a National Institutes of Health predoctoral traineeship, a National Research Service Award, and the California Institute of Technology. I especially thank Mary Arguijo for a superb job in typing and constructing this thesis. Finally, I thank my wife, Debby, for understanding and patience while yet another sequencing gel had to be run, or another paragraph had to be written.

Abstract

A recombinant DNA phage containing a cluster of Drosophila melanogaster tRNA genes has been isolated and analyzed. In situ hybridization localizes this phage to chromosomal region 50AB, a known tRNA site. Nucleotide sequencing of the entire Drosophila tRNA coding region reveals seven tRNA genes spanning 2.5 kb of chromosomal DNA. This cluster is separated from other tRNA regions on the chromosome by at least 2.7 kb on one side, and 9.6 kb on the other. Two tRNA genes are nearly identical and contain intervening sequences in the anticodon loop. These two genes are assigned to be $tRNA^{Leu}$ genes because of significant sequence homology with yeast $tRNA_3^{Leu}$, and secondary structure homology with yeast $tRNA_3^{Leu}$ intervening sequences. In addition, an 8 base sequence (AAAAUCUU) is conserved in the same location in the intervening sequences of Drosophila $tRNA^{Leu}$ genes and a yeast $tRNA_3^{Leu}$ gene. Similar sequences occur in all other tRNAs containing intervening sequences. The remaining five genes are identical $tRNA^{Ile}$ genes, which are identical to a $tRNA^{Ile}$ gene from chromosomal region 42A. The 5' flanking regions are only weakly homologous, but each set of isoacceptors contains short regions of strong homology approximately 20 nucleotides preceding the tRNA coding sequences: GCNTTTTG preceding $tRNA^{Ile}$s; and GANTTTGG preceding $tRNA^{Leu}$s. The genes are irregularly organized on both DNA strands; spacing regions are divergent in sequence and length. Preliminary in vitro transcription experiments with isolated restriction fragments demonstrate that both $tRNA^{Leu}$ genes and at least two of the five $tRNA^{Ile}$ genes are selectively transcribed in vitro by cytoplasmic extracts from HeLa cells.

TABLE OF CONTENTS

Summary

A recombinant DNA phage containing a cluster of Drosophila melanogaster tRNA genes has been isolated and analyzed. In situ hybridization localizes this phage to chromosomal region 50AB, a known tRNA site (Elder, et al., 1980). Nucleotide sequencing of the entire Drosophila tRNA coding region reveals seven tRNA genes spanning 2.5 kb of chromosomal DNA. This cluster is separated from other tRNA regions on the chromosome by at least 2.7 kb on one side, and 9.6 kb on the other. Two tRNA genes are nearly identical and contain intervening sequences in the anticodon loop. These two genes are assigned to be $tRNA^{Leu}$ genes because of significant sequence homology with yeast $tRNA_3^{Leu}$, and secondary structure homology with yeast $tRNA_3^{Leu}$ intervening sequences. In addition, an 8 base sequence (AAAAUCUU) is conserved in the same location in the intervening sequences of Drosophila $tRNA^{Leu}$ genes and a yeast $tRNA_3^{Leu}$ gene. Similar sequences occur in all other tRNAs containing intervening sequences. The remaining five genes are identical $tRNA^{Ile}$ genes, which are also identical to a $tRNA^{Ile}$ gene from chromosomal region 42A (Hovemann, et al., 1980). The 5' flanking regions are only weakly homologous, but each set of isoacceptors contains short regions of strong homology approximately 20 nucleotides preceding the tRNA coding sequences: GCNTTTTG preceding $tRNA^{Ile}$s; and GANTTTGG preceding $tRNA^{Leu}$s. The genes are irregularly organized on both DNA strands; spacing regions are divergent in sequence and length. Preliminary in vitro transcription experiments with isolated

restriction fragments demonstrate that both $tRNA^{Leu}$ genes and at least two of the five $tRNA^{Ile}$ genes are selectively transcribed *in vitro* by cytoplasmic extracts from HeLa cells (Weil, et al., 1979).

Abbreviations used: bp - base pair; kb - kilo base pair; N - any of the nucleotides A, G, C, or T.

Introduction

Several different organizational schemes for eukaryotic tRNA genes have been discovered. For the yeast Saccharomyces cerevisiae, there are approximately 300 tRNA genes, or about 5 genes per tRNA species. These genes are distributed over the yeast genome with little or no clustering (Beckmann, et al., 1977; Olson, et al., 1979). The fruit fly Drosophila melanogaster has an average of 10 genes per tRNA species; these 600 genes occur in clusters of 2 to 15 genes at approximately 50 different chromosomal loci (Steffensen and Wimber, 1971; Weber and Berger, 1976; Elder, et al., 1980). In Drosophila, different tRNA species may be coded within a single gene cluster, while a single tRNA species may be coded in several separate gene clusters (Elder, et al., 1980; Tener et al., 1980; S. Hayashi, et al., 1980). A Drosophila tRNA gene cluster isolated from chromosomal region 42A contains 18 tRNA genes scattered throughout 46 kb of genomic DNA (Hovemann, et al., 1980; Yen and Davidson, 1980). Most of the tRNA species coded by the Drosophila cluster at 42A are repeated within the cluster. These repeated genes are not arranged as tandem repeats, which is the case for Drosophila 5S rRNA genes. In contrast, the african toad Xenopus laevis has about 6,500 tRNA genes (Birnstiel, et al., 1972). Actinomycin-CsCl gradients indicate that most of these genes are clustered on DNA of different G+C content than bulk Xenopus DNA (Clarkson and Kurer, 1976). One Xenopus tRNA gene repeat unit of 3.18 kb is tandemly repeated about 300 times in the genome. This repeat unit contains two $tRNA_1^{Met}$ genes, and potentially codes for

$tRNA^{Phe}$, $tRNA^{Tyr}$, and probably more tRNAs (Clarkson, et al., 1978; Müller and Clarkson, 1980). The reasons for these different schemes of tRNA gene organization are unknown.

Other recent observations about tRNA gene structure have raised interesting questions about their expression. Several yeast tRNA genes and one Xenopus $tRNA^{Tyr}$ gene have been shown to contain intervening sequences: segments of DNA internal to the gene that are not colinear with the tRNA gene product (Goodman, et al., 1977; Knapp, et al., 1978; O'Farrell, et al., 1978; Valenzuela, et al., 1978; Etcheverry, et al., 1979; Ogden, et al., 1979; Venegas, et al., 1979; Müller and Clarkson, 1980). However, the frequency of occurrence of tRNA genes with intervening sequences in eukaryotes is still unknown.

The function of intervening sequences in tRNA genes is also unknown, but their role in the biosynthesis of tRNAs has been partially illuminated. In Xenopus oocytes, the removal of the intervening sequence of injected yeast pre-$tRNA^{Tyr}$ occurs after 3' and 5' end maturations and is the last step prior to transport of the tRNA to the cytoplasm (Melton, et al., 1980). The excision of the intervening sequence and religation of the tRNA half molecules is a two step enzymatic reaction which requires ATP for the second step (Peebles, et al., 1979). It is not known whether these reactions are accomplished by a common enzyme (or set of enzymes), or by separate enzymes for different pre-tRNAs. If a common enzyme is involved, it should recognize some conserved feature, such as

sequence or secondary structure, in the intervening sequence of different precursor tRNAs. Although no one has yet observed any conserved sequences, similar secondary structures have been proposed. The proposed structures usually contain stem loop structures partially within the intervening sequence, and include base pairing of part of the anticodon to the intervening sequence (Knapp, et al., 1978; O'Farrell, et al., 1978; Valenzuela, et al., 1978; Etcheverry, et al., 1979; Ogden, et al., 1979; Venegas, et al., 1979; Müller and Clarkson, 1980).

The genes for small RNAs such as tRNA and 5S rRNA also provide convenient systems for the study of eukaryotic gene transcription. A unique RNA polymerase, polymerase III, transcribes these RNAs, as well as small nuclear and viral RNAs (Reviewed by Roeder, 1976). In vivo and in vitro RNA polymerase III transcription systems accurately and specifically transcribe cloned tRNA and 5S rRNA genes. Xenopus oocytes accurately transcribe injected genes and process homologous and heterologous tRNAs (Brown and Gurdon, 1977, 1978; Kressman, et al., 1977; Melton, et al., 1980). Xenopus RNA polymerase III is also active in nuclear extracts obtained from oocyte germinal vesicles (Birkenmeier, et al., 1978; Schmidt et al., 1978). R. G. Roeder and coworkers (Weil, et al., 1979) have developed another in vitro polymerase III transcription system from cytoplasmic extracts of human tissue culture cells. The cytoplasmic extracts accurately transcribe exogenous 5S and small viral RNA genes. We report that this system also transcribes Drosophila tRNA genes.

Ultraviolet transcriptional mapping and numerous transcription studies indicate that eukaryotic tRNAs and 5S rRNAs are transcribed as small, independent units (Feldmann, 1977; Brown and Gurdon, 1977, 1978; Schmidt, et al., 1978; Garber and Gage, 1979).

The transcription of Xenopus 5S rRNA genes is controlled by an intriguing mechanism: a small protein binds to an intragenic control region, allowing transcription initiation (Sakonju, et al., 1980; Bogenhagen, et al., 1980; Engelke, et al., 1980). Sequences flanking the 5' end of the gene do not affect proper initiation, but sequences flanking the 3' end are required for proper termination.

An intragenic control region has also been found for Drosophila $tRNA^{Lys}$ genes (DeFranco, et al., 1980). However, in contrast to the Xenopus results, the 5' flanking sequences of one $tRNA_2^{Lys}$ gene negatively modulated transcription of the gene.

The comparative study of the 5' flanking sequences of RNA polymerase III transcripts may yet yield insights into the control of transcription of these genes.

We have continued our earlier studies (Yen, et al., 1977) of Drosophila tRNA gene organization. We report here the isolation, sequence analysis, and in vitro transcription of a cloned Drosophila tRNA gene cluster. Five regions in the cluster potentially code for $tRNA^{Ile}$, and two regions potentially code for $tRNA^{Leu}$. Both $tRNA^{Leu}$ genes have intervening sequences and share significant homologies with yeast $tRNA_3^{Leu}$ genes. The 5' flanking sequences of the genes contain weak homologies, and several genes are accurately transcribed by an in vitro extract.

Results

Isolation and sequencing of a tRNA cluster

A recombinant DNA phage containing a cluster of *Drosophila* tRNA genes was obtained by screening a recombinant library by the plaque filter hybridization method of Benton and Davis (1977). The recombinant library was constructed by cloning partial Eco RI digests of *Drosophila* nuclear DNA in the modified λ cloning vector Charon 4 (Yen, et al., 1979; Davidson, et al., 1980). Previously isolated recombinant plasmids, which contained one or more *Drosophila* tRNA genes cloned in the vector Col El, were labeled with ^{32}P and used as hybridization probes for the screen (P. Yen, unpublished). 14 strongly hybridizing recombinant phage were isolated from this screen. DNA from each phage was bound to nitrocellulose filters and hybridized to ^{3}H labeled tRNA isolated from cultured Schneider *Drosophila* cells. The recombinant phage giving the strongest hybridization signal was estimated to contain 5-10 tRNA genes. *In situ* hybridization localized this phage to region 50AB of chromosome 2R (data not shown), and it was therefore named Ch4:Dm50AB.

The restriction enzyme mapping and Southern hybridization analysis of Ch4:Dm50AB using ^{32}P labeled pupal tRNA as a probe indicated that all tRNA coding regions were located on a 3.4 kb stretch of DNA between a Sal I and an Eco RI restriction site (Fig. 1). In addition, four separate restriction fragments within the 3.4 kb Sal I-Eco RI fragment hybridized to the tRNA probe, indicating that multiple tRNA genes were located between the Sal I and Eco RI sites.

The Sal I-Bam HI restriction fragment containing the entire tRNA hybridizing region was subcloned into the plasmid vector pBR322, resulting in the recombinant plasmid p50AB. Additional restriction enzymes were used to further map the tRNA coding region between the Sal I and Eco RI sites of plasmid p50AB. The subcloning and restriction mapping are shown schematically in Figure 2.

The entire nucleotide sequence of the tRNA coding region was determined by the method of Maxam and Gilbert (1980). A summary of the sequence runs is depicted in Figure 2. 55% of the sequence was obtained by sequencing both strands of a region or by independently sequencing the same strand. When comparing the sequence from opposite strands, we observed a discrepancy between the two readings of about 2%. Therefore, we expect approximately a 2% error rate in those regions that were sequenced only once. However, most tRNA gene regions were sequenced at least twice. The sequence had no discrepancies with the restriction map of the clone, with the exception of several predicted Taq I sites which were not cleaved by Taq I. At all of these sites, however, the Taq I recognition sequence overlapped with an Mbo I recognition sequence. This particular combination of overlapping sites has been observed to confer resistance to Taq I cleavage, probably due to A methylation within the Taq I site when the plasmid is grown in E. coli (Müller and Clarkson, 1980). Thus, we find no unexplainable discrepancies between the restriction enzyme map and the sequence.

The complete nucleotide sequence between the Sal I and Eco RI sites is presented in Appendix I.

Identification of tRNA genes

The strongest conservation of sequence within eukaryotic cytoplasmic tRNAs is in the Tψ loop region (Clark, 1978; Sprinzl, et al., 1980). With the aid of a computer program (Staden, 1977), we searched the nucleotide sequence of p50AB for possible Tψ loop regions. The sequence for all candidate tRNA regions were then folded into a cloverleaf secondary structure and checked for tRNA conserved sequences and secondary structure.

Seven potential tRNA coding regions were discovered. These coding regions, their flanking sequences, and their cloverleaf structures are shown in Figure 3. Five of the genes code for an identical tRNA sequence which contains the isoleucine anticodon AAU. An identical $tRNA^{Ile}$ coding sequence has been previously sequenced on a recombinant plasmid from region 42A of the Drosophila chromosome (Hovemann, et al., 1980). The remaining two tRNA coding regions are quite different from the $tRNA^{Ile}$ coding sequences, but are nearly identical with each other, and unexpectedly contain intervening sequences. As in all other tRNA genes containing intervening sequences, the location of the intervening sequences is within the anticodon loop. In our case, this location causes some ambiguity in the identification of the tRNAs due to the possibility that the splice junction may be within the anticodon of the mature tRNA. Knapp, et al. (1979), however, have found that there is a conserved site for the splice junction of all yeast tRNAs containing intervening sequences that have been analyzed. The conserved splice junction is just 3' of the purine nucleotide which is 3' to the anticodon. Following this rule for the position of the splice

junction, we have tentatively identified the Drosophila tRNA genes as leucine tRNA genes, containing the anticodon CAA. The differences between the sequences for the $tRNA^{Leu-a}$ gene and the $tRNA^{Leu-b}$ gene lie entirely in the intervening sequences, and consist of one T→G transversion and an insertion of seven nucleotides (Figure 3).

Sequence organization of the tRNA genes and spacers

Four $tRNA^{Ile}$ genes are located on the same strand as the $tRNA^{Leu-b}$ gene, corresponding to rightward transcription in Figure 1. The $tRNA^{Ile-a}$ and $tRNA^{Leu-a}$ genes are located on the opposite strand, corresponding to leftward transcription, divergent from the other five genes.

The spacing regions between the genes are variable in length (339; 165; 230; 409; 237; and 491 bases, from left to right in Figure 2). None of the tRNA coding regions includes a genomically coded CCA 3' end. Initial inspection of the sequences near the tRNA genes revealed that the preceding and trailing sequences diverge immediately beyond the mature tRNA coding sequences. Thus, there are no repetitive units within this cluster except for the mature tRNA coding sequences themselves.

Examination of the restriction map of the parent Ch4:Dm50AB indicated that the tRNA gene cluster is separated from any other tRNA coding regions by 9.6 kb and 2.7 kb on the Drosophila genome. There are no tandemly repeated units within the Drosophila DNA insert of Ch4:Dm50AB. If the insert is part of a tandemly repeated array, then the repeat unit would have a minimum length of 15.6 kb, the length of the Ch4:Dm50AB insert.

In vitro transcription

The tRNA gene cluster was transcribed *in vitro* with HeLa cell cytoplasmic extracts containing RNA polymerase III activity (Weil, et al., 1979). The results of these experiments are shown in Figure 4. In all of the transcription experiments, we used single stranded ^{32}P labeled DNAs as convenient length standards to approximate RNA lengths. All acrylamide gels were 8M urea and and run at 60°C, with heat denaturation of samples in 80% formamide prior to loading. Under these conditions we find that the main band of *Drosophila* tRNA is located between 72 and 80 base single strand DNA markers. Since this is a reasonable size for tRNA, we believe that the size estimations are fair approximations. However, molecules that have an unusually high degree of secondary structure, or very short molecules, may migrate at anomalous lengths. In the results reported below, we find that RNA oligonucleotides comigrate with DNA oligonucleotides of approximately 10% greater length in a 25% acrylamide, 8 M urea gel. A similar case has been found for 20% acrylamide, 8 M urea gels (R. Ogden, personal communication).

The *in vitro* transcription of supercoiled plasmid p50AB DNA produced several bands: two large bands of approximate length 145 and 137 bases; a heavy smear from 95 to 85 bases; and a ladder of smaller bands from 64 to 40 bases, including strong bands at 55, 51, 47, 44 and 40 bases (Figure 4A). In our hands, there was considerable background at mature tRNA size in the α-^{32}P-UTP label *in vitro* transcription reactions. We presume that this is due to

amination of α-^{32}P-UTP to α-^{32}P-CTP by the extract, followed by terminal addition of (^{32}P)CCA to endogenous tRNAs in the extract. There was little background from the α-^{32}P-GTP label reactions.

In order to determine the origins of the various transcripts, we transcribed isolated restriction fragments containing tRNA coding regions (Figure 4B). The transcription of a 588 bp Alu I fragment containing the $tRNA^{Leu-a}$ gene produced the 137 base band, a ladder of bands from 64 to 50 bases, and moderate bands at 44 and 40 bases. The transcription of a 644 bp Alu I fragment containing the $tRNA^{Leu-b}$ gene produced a similar pattern of bands, except the 137 base band was replaced by a 145 base band, and the 40 base band was replaced by a 47 base band. The transcription product of a 618 bp Hha I fragment containing the $tRNA^{Ile-a}$ gene was a ladder of bands from 91 to 79 bases, while the transcription products of the 1773 bp Hha I fragment containing the $tRNA^{Ile-b}$, $tRNA^{Ile-c}$, $tRNA^{Ile-d}$, and $tRNA^{Ile-e}$ genes were two sets of bands centered at approximately 95 and 88 bases. However, a 674 bp Pvu II fragment containing the $tRNA^{Ile-d}$ gene transcribed extremely weakly, and a 169 bp Pvu II-Hind III fragment containing the $tRNA^{Ile-e}$ gene did not transcribe at all. Although the $tRNA^{Ile}$ transcriptions produced a moderate band of approximate length 79 bases, none of the $tRNA^{Leu}$ transcriptions produced any significant bands corresponding to mature tRNA length.

Therefore we have identified the <u>in vitro</u> transcription bands from the tRNA plasmid p50AB: the 137 base transcript is from the $tRNA^{Leu-a}$ gene region; the 145 base transcript is from the $tRNA^{Leu-b}$

gene region; the 95-85 base transcripts are from the $tRNA^{Ile}$ gene regions; and the ladder of bands from 64 to 40 bases is from the $tRNA^{Leu}$ gene regions.

The 137 and 145 base bands are precursors of smaller bands

We surmised that the 137 and 145 base transcripts from the $tRNA^{Leu}$ gene regions were primary transcripts containing the $tRNA^{Leu-a}$ and $tRNA^{Leu-b}$ sequences, respectively. In order to determine if the 137 and 145 base transcripts were precursors of smaller bands appearing during *in vitro* transcription, we isolated the ^{32}P labeled 137 and 145 base RNAs and chased them in an unlabeled *in vitro* transcription reaction. Over a span of 60 min., both the 137 and 145 base bands were chased into sets of small bands: the 137 base transcript produced bands at 40, 44, and a weak ladder from 50 to 56; the 145 base transcript produced bands at 44, 47, and a weak ladder from 50 to 56 (Figure 5, compare to Figure 4B). Therefore, at least some of the smaller bands produced during the transcription of the $tRNA^{Leu}$ genes are processing or degradation products from the large transcripts.

RNAase T1 and RNAase A oligonucleotides of in vitro transcripts are predicted by the DNA sequence

To prove that the 137 and 145 base bands were due to transcription of the $tRNA^{Leu}$ genes, and not some other transcription site on the restriction fragments, we digested the transcripts with RNAase T1 and RNAase A and analyzed the products. First, the isolated ^{32}P labeled 137 and 145 base RNAs were digested with RNAase T1

and the products were run on a denaturing 25% acrylamide, 8M urea gel. The lengths of the T1 oligonucleotides were compared to the lengths predicted from the DNA sequences of the $tRNA^{Leu}$ genes (Figure 6 and Table 1). The largest T1 oligonucleotide predicted from either transcript is located within the intervening sequence, and has a length of 20 bases for $tRNA^{Leu-a}$ and 16 bases for $tRNA^{Leu-b}$ Each of these T1 oligonucleotides were present as predicted, comigrating with single strand DNA markers of 22 and 18 bases, respectively. In addition to these two large characteristic oligonucleotides, the almost identical sequence of the two $tRNA^{Leu}$ genes predicted several identical T1 oligonucleotides, which were observed in all cases (Table 1). Also, the $tRNA^{Leu}$ gene sequences predicted some cases of differential α-^{32}P-GTP or α-^{32}P-UTP label, which were also observed.

To further establish the identity of the T1 oligonucleotides, several characteristic oligonucleotides were isolated from the 25% acrylamide gel, digested with RNAase A, and electrophoresed on DEAE paper at pH 3.5. The predicted patterns for the 20 base and 16 base T1 oligonucleotides from $tRNA^{Leu-b}$ and $tRNA^{Leu-a}$ transcripts include no RNAase A products greater than 2 bases except for an A_4U, which is present in each. As predicted, we found no RNAase A oligos greater than 2 bases except for a band migrating near the origin, characteristic of $A \geq_4 N$ (data not shown). In addition, the RNAase A digest of the 13 base T1 oligonucleotides from both $tRNA^{Leu-a}$ and $tRNA^{Leu-b}$ gene transcripts yielded no oligonucleotide greater than 2 bases except for an AAU oligonucleotide, also as predicted from the DNA sequence (data not shown).

Discussion

The $tRNA^{Leu}$ genes from p50AB

There are two arguments that the mature tRNAs from the two tRNA genes containing intervening sequences are identical leucine tRNAs containing the anticodon CAA. First, as mentioned above, previous studies of five yeast pre-tRNAs containing intervening sequences have demonstrated a conserved position for the splice junction: it is always located between the two nucleotides just to the 3' side of the anticodon (Knapp, et al., 1979). To comply with the rules for tRNA sequence and secondary structure (Clark, 1978), the intervening sequence of the p50AB $tRNA^{Leu}$ genes must lie within the anticodon loop. This location is consistent with the conserved site found for yeast tRNA intervening sequences.

Secondly, there are significant homologies in sequence and secondary structure between yeast $tRNA_3^{Leu}$ genes (Venegas, et al., 1979; Kang, et al., 1980; Johnson, et al., 1980) and the proposed Drosophila $tRNA^{Leu}$ genes. Comparisons of the secondary structure of the yeast pre-$tRNA_3^{Leu}$s and the proposed Drosophila pre-$tRNA^{Leu}$s were made after folding the sequences into cloverleaf form (Figure 7) and minimizing the free energy of all stem loop structures (Tinoco, et al., 1973; Borer, et al., 1974). The secondary structures involving the intervening sequences are almost identical (Figure 7). There are two stem loop structures involving the intervening sequences: the first has a stem of 5 bp in both Drosophila and yeast, while the second immediately follows the first with a stem of 6 bp. The sizes of the

loops differ slightly for the Drosophila and yeast precursors, from 7 to 5 bases for the first loop, and from 4 to 5 bases for the second. The predicted free energies of these two stem loop structures are -3.5 kcal and -4.0 kcal for the yeast pre-$tRNA_3^{Leu}$s, and -1.0 kcal and -3.7 kcal for the proposed Drosophila pre-$tRNA^{Leu}$s. The CAA leucine anticodon is completely base paired to the yeast intervening sequence, and the last two anticodon bases are base paired to the Drosophila intervening sequence.

In addition to the conservation of secondary structures, there is also strong sequence homology between the Drosophila and yeast $tRNA_3^{Leu}$ genes (Figure 8). Most of the 65% sequence homology between these genes occurs in the dihydrouridine stem and loop, in the mature tRNA anticodon stem and loop, and in the Tψ loop. Of the 47 nucleotides in these regions, 43 are conserved between the Drosophila and yeast genes. The 91% homology in these regions is much stronger than observed when comparing the Drosophila $tRNA^{Leu}$ genes with other tRNAs: 74% with yeast $tRNA^{Tyr}$; 70% with yeast $tRNA^{Phe}$; 53% with Drosophila $tRNA^{Asn}$; and 51% with Drosophila $tRNA^{Ile}$. These striking homologies suggest that the yeast $tRNA_3^{Leu}$ genes and our Drosophila $tRNA^{Leu}$ genes have a common ancestor.

The intervening sequences of the yeast and Drosophila $tRNA^{Leu}$ genes contain two regions of strong sequence homology (Figures 7,8).

One 8 base region, AAAAUCUU, in the Drosophila $tRNA^{Leu}$ intervening sequences is exactly repeated in approximately the same location in one yeast pre-$tRNA_3^{Leu}$ (Venegas, et al., 1979). An almost identical sequence, AAUAUCUU, occurs at the same position in the other yeast pre-$tRNA_3^{Leu}$ (Kang, et al., 1980; Johnson, et al., 1980). The homology originates in the first loop of the intervening sequence and continues into the stem which base pairs to the anticodon (Figure 7). A second homology of 5 bases occurs mainly within the second loop of the intervening sequences. The sequence in this region is UUAAC for one yeast pre-$tRNA_3^{Leu}$, and UGAAC for the other yeast and Drosophila $tRNA^{Leu}$ genes. The probabilities of finding exact homologies in the same approximate location of two random sequences of intervening sequence length are .12 for 5 bases and .002 for 8 bases.

The conservation of an 8 base region between the $tRNA^{Leu}$ intervening sequences of very different organisms is an intriguing observation, and suggests that this sequence may be conserved in other tRNA genes containing intervening sequences. We have scanned published tRNA intervening sequences and have found similar sequences starting within all known tRNA intervening sequences (Table 2). Of six tRNA intervening sequences other than $tRNA^{Leu}$, three had a homology of 7 nucleotides with the sequence AAAAUCUU, two had a homology of 6, and one had a homology of 5.

This family of conserved sequences starting within tRNA intervening sequences has additional similarities. The last 2 nucleotides, and usually more, are always base paired in a predicted stable secondary

structure. The first few nucleotides are always located in a loop region. The last nucleotide, and often more, is base paired to an anticodon nucleotide except for the two $tRNA^{Tyr}$ precursors. These similarities in sequence and secondary structure indicate that these sequences are strongly related.

The homologies of this family of conserved sequences suggest a functional significance. One possible functional role might be the control of transcription of the tRNA gene (Johnson, et al., 1980). If the control of transcription of tRNA genes is similar to that for 5S rRNA genes (Bogenhagen, et al., 1980; Sakonju, et al., 1980; Engelke, et al., 1980), then the intervening sequences would lie in the transcription initiation control region. Johnson, et al. (1980) have found that an insertion in the second stem loop structure of yeast pre-$tRNA_3^{Leu}$ did not affect *in vitro* transcription and processing of the gene. However, the conserved sequence lies within the first loop structure; alterations in the second loop structure might not affect transcriptional control exerted by the conserved sequence.

There are, however, some observations that the intervening sequence transcriptional control hypothesis does not explain. Many tRNA genes do not have intervening sequences. Also, many of these tRNAs have no sequence similarity to the 8 base conserved sequences (*Drosophila* $tRNA^{Ile}$, for example). Thus, neither an intervening sequence nor the 8 base conserved sequence is required for the transcription of all tRNA genes.

Another possible function of these conserved sequences is to serve as a common recognition site for the enzyme or enzymes responsible for the removal of the intervening sequence. It is not yet known whether a single enzyme or different enzymes are required for removal of the intervening sequences of different tRNAs, but the recent isolation of yeast mutants which only affect tRNA intervening sequence metabolism (Hopper, et al., 1980) implies that a common enzyme may be involved. Our observations are consistent with a common enzyme involved in tRNA intervening sequence excision: the conserved family of sequences within the intervening sequence may provide a common recognition site for this enzyme.

Homologies of Drosophila tRNA gene flanking regions

There are imperfect homologies among the 5' flanking regions of the $tRNA^{Ile}$ genes. With no gaps or loopouts allowed, the best alignment of the 50 nucleotides preceding the $tRNA^{Ile-b}$ and $tRNA^{Ile-e}$ genes produces a 62% sequence homology (Figure 9). Homologies in this region among the other $tRNA^{Ile}$ genes are 36% to 48%. The 5' flanking region of the $tRNA^{Ile}$ gene from chromosomal region 42A (Hovemann, et al., 1980) has greatest homology (44%) with the sequence preceding the $tRNA^{Ile-b}$ gene. Although these homologies are weak, they are greater than homologies obtained from comparison of random sequences. These results suggest that the repeated $tRNA^{Ile}$ genes and their spacers may have been created by a gene duplication of an ancestral gene most closely related to the $tRNA^{Ile-b}$ gene.

A small region of strong homology precedes the p50AB $tRNA^{Ile}$ genes, starting 23 or 24 nucleotides from the 5' end of the $tRNA^{Ile}$ coding region. A consensus sequence derived from the nontranscribed strand of this region, 5' GCNTTTTG 3', occurs in the 5' flanking sequences of the $tRNA^{Ile-b}$ and $tRNA^{Ile-e}$ genes, while sequences which match 6 of the 7 consensus nucleotides precede the other $tRNA^{Ile}$ genes (Figure 3). The sequences from the $tRNA^{Ile}$ gene from chromosomal region 42A contain a run of 8 T nucleotides starting 23 nucleotides from the tRNA coding region, but otherwise are not homologous to the consensus sequence preceding p50AB $tRNA^{Ile}$ genes. The 5' flanking sequences of the $tRNA^{Leu}$ genes also contain a small conserved sequence, 5' GANTTTGG 3', preceding the $tRNA^{Leu-a}$ and $tRNA^{Leu-b}$ coding regions by 12 and 21 nucleotides, respectively. Similar small regions of imperfect homology have been found preceding repeated Drosophila $tRNA_2^{Lys}$ and $tRNA^{Arg}$ genes. Hovemann, et al. (1980) have determined a consensus sequence from the 5' flanking sequences of the Drosophila $tRNA_2^{Lys}$ genes: 5' GGCAGTTTTTA 3' (Hovemann, et al., 1980; Yen and Davidson, 1980). It is interesting that the consensus sequences are G and T rich; they also contain runs of 3 or more T nucleotides.

Whether the conservation of these small homologous regions preceding repeated tRNA genes is due to function or some other reason remains unclear. However, there is some evidence that 5' flanking sequences may modulate transcription of Drosophila $tRNA_2^{Lys}$ genes by Xenopus RNA polymerase III (DeFranco, et al., 1980).

The sequences flanking the 3' end of all the p50AB tRNA genes contain strings of 6 or more T nucleotides in the nontranscribed strand, starting within 10 nucleotides of the tRNA coding region. Since all RNA polymerase III transcripts that have been studied terminate within runs of 4 or more T nucleotides in the nontranscribed strand (Silverman, et al., 1979; Korn and Brown, 1978; Garber and Gage, 1979), the stretches of T nucleotides trailing tRNA genes are probably termination sites.

In vitro transcription

Preliminary in vitro transcription experiments suggest that the $tRNA^{Leu}$ genes and several $tRNA^{Ile}$ genes are selectively transcribed. Therefore, these tRNA genes appear to contain all information required for accurate initiation and termination of transcription by RNA polymerase III in HeLa cell extracts. We were not surprised that the HeLa system transcribed Drosophila tRNA genes because human cell extracts accurately transcribe Xenopus 5S genes (Weil, et al., 1979). Also, Xenopus RNA polymerase III systems accurately transcribe a number of heterologous tRNA genes: Drosophila $tRNA^{Arg}$ (Silverman, et al., 1979); Bombyx mori $tRNA_2^{Ala}$ (Garber and Gage, 1980); and yeast tRNAs (Ogden, et al., 1979; Johnson, et al., 1979). Clearly, the RNA polymerase III transcriptional apparatus is well conserved in eukaryotes.

When labeled primary transcripts of the $tRNA^{Leu}$ genes are chased with unlabeled in vitro transcription extracts, the transcripts are partially processed, producing small bands of 56 to 40 bases. One of

the small processed fragments that we might expect to observe would be the excised intervening sequence from each $tRNA^{Leu}$ transcript, predicted to be 38 bases for $tRNA^{Leu-a}$ and 45 bases for $tRNA^{Leu-b}$. Indeed we find candidate bands comigrating with single strand DNAs of 40 and 47 nucleotides, respectively (Figure 4). Some preliminary experiments indicate that the $tRNA^{Leu-a}$ 40 base band contains the RNAase T1 oligonucleotide characteristic of the intervening sequence (data not shown). However, we have not yet positively identified the 40 base fragment.

Other fragments expected from processing of $tRNA^{Leu}$ transcripts would be the 5' and 3' half tRNAs left after excision of the intervening sequence, and the mature tRNA.

A preliminary experiment indicates that a band from $tRNA^{Leu-a}$ transcription, which comigrates with 44 base single strand DNA, contains RNAase T1 oligonucleotides specific only for the 5' half of $tRNA^{Leu}$. However, we have not identified any small transcription bands which originate only from the 3' half of the tRNA, and we have not found any significant bands corresponding to mature $tRNA^{Leu}$. The absence of mature $tRNA^{Leu}$ implies that at least some of the tRNA processing enzymes are not present or active in the *in vitro* transcription extracts.

Drosophila tRNA gene organization

We do not know whether the tRNA genes from Ch4:Dm50AB are the only tRNA genes in chromosomal region 50AB; other tRNA genes may be

located on the chromosome near this cluster (but at least 2.7 kb away), which would make the tRNA genes reported here a sub-cluster of a larger region of tRNA genes. Nevertheless, the overall organization of the tRNA gene cluster in Ch4:Dm50AB is consistent with the results of previous investigations (Yen, et al., 1977; Tener, et al., 1980; Elder, et al., 1980; Hovemann, et al., 1980; Yen and Davidson, 1980). Drosophila tRNAs are organized in clusters of varying sizes. A cluster at chromosomal region 42A spans 46 kb and contains 18 tRNA coding regions: 8 $tRNA^{Asn}$, 4 $tRNA^{Arg}$, 5 $tRNA^{Lys}$, and 1 $tRNA^{Ile}$. The cluster reported here spans 2.5 kb and includes 7 tRNA coding regions: 5 $tRNA^{Ile}$ and 2 $tRNA^{Leu}$. Although individual tRNA genes are often repeated within a cluster, the genes are irregularly spaced and located on both chromosomal DNA strands. The spacer regions are divergent in nucleotide sequence, but small homologies exist preceding repeated genes.

The reasons for this type of clustering, as opposed to the tandemly repeated transcriptional units of Drosophila 5S rRNA genes (Hershey, et al., 1977; Artavanis-Tsakonas, et al., 1977), are unclear. Perhaps the relatively small number of genes for a specific tRNA is required to be more constant throughout evolution than is the number of 5S rRNA genes. The tandem repetitive organization of 5S rRNA genes allows rapid expansion or contraction of the number of genes by unequal crossover events (Smith, 1973, 1976). Because of their irregular sequence organizations, tRNA genes would not be as susceptible as 5S genes to unequal crossover, reducing the rapid

contraction or expansion of repeated gene numbers by recombination events.

Experimental Procedures

Screening the recombinant library

The construction and analysis of the recombinant library has been described previously (Yen, et al., 1979; Davidson, et al., 1980). Partial Eco RI digested nuclear DNA from pupae of *Drosophila melanogaster* strain Canton S was size selected on sucrose gradients and cloned into the λ vector Charon 4 (Blattner, et al., 1977). Insert sizes range from 10-21 kb, and Cot analysis indicates that the library is representative of 95 ± 5% of the single copy *Drosophila* genome. Dr. Pauline Yen combined three recombinant plasmids containing one or more *Drosophila* tRNA genes cloned in the vector Col El (Yen, et al., 1977), labeled them by nick translation (Maniatis, et al., 1975), and used them as hybridization probes to screen the library (Benton and Davis, 1977; Maniatis, et al., 1978).

Isolation and labeling of tRNA

The isolation of Drosophila pupal tRNA and ^{3}H labeled *Drosophila* Schneider cell tRNA has been described (Yen, et al., 1977). *Drosophila* pupal tRNA was labeled with ^{32}P by two different methods, each described in detail by Yen and Davidson (1980): tRNA 5' ends were labeled by treatment with T4 polynucleotide kinase and γ-^{32}P-ATP (Donis-Keller, et al., 1977); tRNA 3' ends were labeled by treatment with T4 RNA ligase and ^{32}P-pCp (Barrio, et al., 1978).

Gel electrophoresis, Southern hybridization, and restriction mapping

Recombinant phage DNA was digested singly or in combination by restriction enzymes, and the products were separated by 0.7% or 1.5% agarose gel electrophoresis in E buffer [40 mM Tris, 5 mM NaOAc, 1 mM EDTA (pH 7.4)]. Restriction fragments hybridizing to tRNA were detected by transfer of the fragments to a nicrocellulose filter (Southern, 1975), and hybridization to 0.1 µg/ml ^{32}P-tRNA at 65°C overnight in 1 M NaCl, 50 mM Tris (pH 7.5), 10 mM EDTA, 0.1% SDS, 0.2% bovine serum albumin, 0.2% PVP-40, and 0.2% Ficoll. A restriction map of the phage DNA was deduced by analysis of restriction fragment lengths on agarose gels and the ability of restriction fragments to hybridize to ^{32}P-tRNA probes. A restriction map of the plasmid p50AB was deduced by analysis of restriction fragment lengths on 5% acrylamide gels in TBE buffer [90 mM Tris, 90 mM HBO_4, 2.5 mM EDTA (pH 8.3)], and by partial digestion of end labeled restriction fragments (Smith and Birnstiel, 1976).

In situ hybridization

Dr. Pauline Yen prepared ^{3}H-cRNA from Ch4:Dm50AB and then hybridized it *in situ* to squashed *Drosophila* larval chromosomes as described by Yen, et al. (1977).

Construction of plasmid p50AB

1.5 µg Ch4:Dm50AB and 0.45 µg plasmid pBR322 DNA (Bolivar, et al., 1977) were mixed and digested with Sal I and Bam HI restriction enzymes. The resulting restriction fragments were ligated with T4 DNA ligase

(Mertz and Davis, 1972; Cohen, et al., 1973) and the ligation mixture was used to transform E. coli strain HB101. Transformants containing Drosophila tRNA genes were selected by colony hybridization (Grunstein and Hogness, 1975) to ^{32}P-tRNA probe. Closed circulur DNA from one positive colony, p50AB, was prepared from a bacterial lysate of chloramphenicol-amplified culture (Clewell and Helinski, 1972) by ethidium bromide-cesium chloride equilibrium sedimentation (Radloff, et al., 1967).

Isolation, labeling, sequencing of restriction fragments

Restriction fragments were isolated and labeled at 5' ends with T4 polynucleotide kinase and γ-^{32}P-ATP as described by Maxam and Gilbert (1980). Fragments were labeled at 3' ends with E. coli DNA polymerase (Klenow fragment): 1 µg of DNA was treated with 0.2 - 1 units DNA polymerase (Klenow) in 30 µl of 5 mM dithiothreitol, 6.6 mM NaCl, 6.6 mM $MgCl_2$, and 6.6 mM Tris (pH 7.5) with 1 µM each of ^{32}P-dATP, ^{32}P-dCTP, ^{32}P-dGTP, and ^{32}P-dTTP (2000 Ci/mmole) at 37°C for 30 min. End labeled fragments were sequenced by the method of Maxam and Gilbert, using reactions G, G+A, A>C, C+T, and C.

Computer programs

Minor modifications of the programs described by Staden (1977, 1978) were used with a Digital Equipment VAX-11 computer to analyze DNA sequences.

In vitro transcription and RNA analysis

Extracts of HeLa cells, prepared by the method of Weil, et al. (1979) by Dr. Ed Fritsch, were used for *in vitro* transcription experiments. The products of *in vitro* transcription were run on 8% denaturing acrylamide gels (Maxam and Gilbert, 1980), and ^{32}P-RNA bands were eluted from the gels as described for DNA fragments. Unlabeled *E. coli* tRNA was added to the eluted ^{32}P-RNA, ethanol precipitated, and resuspended in 5 µl RNAase T1 (10 units) in water. After incubation at 37°C for 30 min, the ^{32}P-RNA was added to an equal volume of 80% formamide loading buffer, heated to 90°C for 5 min, chilled in ice, and loaded onto a 25% acrylamide, 8 M urea gel (Maxam and Gilbert, 1980). RNA bands from this gel were eluted as before and digested in 20 µl of 0.1 mg/ml RNAase A at 37°C for 30 min, then loaded and run on DEAE paper at pH 3.5 (Brownlee, 1972).

Containment of recombinant DNA organisms

All recombinant DNA containing organisms were handled in P2, EK1 conditions in accordance with the NIH Guidelines.

Acknowledgements

Our collaborators, Dr. Pauline Yen, Dr. N. Davis Hershey, and Dr. Kenshi Hayashi provided many helpful discussions. We thank Dr. Yen for the original gift of Ch4:Dm50AB and for assistance in its localization by *in situ* hybridization. We gratefully acknowledge Dr. Ed Fritsch for gifts of *in vitro* transcription extracts and instruction in their use. We thank Dr. Rodger Staden for making available his computer programs for DNA sequence analysis, and Dr. Pamela Mellon for instruction in DEAE paper electrophoresis. RR was a recipient of a National Institutes of Health predoctoral traineeship and National Research Service Award 5 T32 GM 07616-02 from the National Institute of General Medical Science. Additional funding was supplied by a grant from the National Institutes of Health.

References

Artavanis-Tsakonas, S., Schedl, P., Tschudi, C., Pirotta, V., Steward, R., and Gehring, W.J. (1977). The 5S genes of Drosophila melanogaster. Cell 12, 1057-1067.

Barrio, J., del Carmen G. Barrio, M., Leonard, N.J., England, T.E., and Uhlenbeck, O.C. (1978). Synthesis of modified nucleoside 3',5'-Bisphosphates and their incorporation into oligoribonucleotides with T4 RNA ligase. Biochemistry 17, 2077-2081.

Beckmann, J.S., Johnson, P.F., and Abelson, J. (1977). Cloning of yeast transfer RNA genes in Escherichia coli. Science 196, 205-208.

Benton, W.D. and Davis, R.W. (1977). Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196, 180-182.

Birkenmeier, E.H., Brown, D.D. and Jordan, E. (1978). A nuclear extract of Xenopus laevis oocytes that accurately transcribes 5S RNA genes. Cell 15, 1077-1086.

Birnstiel, M.L., Sells, B.H. and Purdom, I.F. (1972). Kinetic complexity of RNA molecules. J. Mol. Biol. 63, 21-39.

Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L. and Smithies, O. (1977). Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. Science 196, 161-169.

Bogenhagen, D.F., Sakonju, S. and Brown, D.D. (1980). A control region in the center of the 5S RNA gene directs specific initiation of transcription: II. The 3' border of the region. Cell 19, 27-35.

Bolivar, F., Rodriguez, R., Greene, P.J., Betlach, M., Heyneker, H.L., Boyer, H.W., Crosa, J. and Falkow, S. (1977). Construction and characterization of new cloning vehicles. Gene 2, 95-113.

Borer, P.N., Dengler, B., Tinoco, I., Jr. and Uhlenbeck, O.C. (1974). Stability of ribonucleic acid double stranded helices. J. Mol. Biol. 86, 843-853.

Brown, D.D. and Gurdon, J.B. (1977). High-fidelity transcription of 5S DNA injected into Xenopus oocytes. Proc. Nat. Acad. Sci. USA 74 2064-2068.

Brown, D.D. and Gurdon, J.B. (1978). Cloned single repeating units of 5S DNA direct accurate transcription of 5S RNA when injected into Xenopus oocytes. Proc. Nat. Acad. Sci. USA 75, 2849-2853.

Brownlee, G.G. (1972). Determination of sequences in RNA. In Laboratory Techniques in Biochemistry and Molecular Biology, T.S. Work and E. Work, eds. (Elsevier, New York), pp. 67-99.

Clark, B.F.C. (1978). General features and implications of primary, secondary and tertiary structure. In Transfer RNA, S. Altman, ed., (MIT Press, Cambridge, Massachusetts), pp. 14-47.

Clarkson, S.G. and Kurer, V. (1976). Isolation and some properties of DNA coding for $tRNA_1^{Met}$ from Xenopus laevis. Cell 8, 183-195.

Clarkson, S.G., Kurer, V. and Smith, H.O. (1978). Sequence organization of a cloned $tDNA_1^{Met}$ fragment from Xenopus laevis. Cell 14, 713-724.

Clewell, D.B. and Helinski, D.R. (1972). Effect of growth conditions of the formation of the relaxation complex of supercoiled Col El deoxyribonucleic acid and protein in Escherichia coli. J. Bacter. 110, 1135-1146.

Cohen, S.N., Chang, A.C.Y, Boyer, H.W. and Helling, R.B. (1973). Construction of biologically functional bacterial plasmids in vitro. Proc. Nat. Acad. Sci. USA 70, 3240-3244.

Davidson, N., Fyrberg, E.A., Hershey, N.D., Kindle, K., Robinson, R.R., Sodja, A. and Yen, P. (1980). Recombinant DNA studies of DNA sequence organization around actin and tRNA genes of Drosophila melanogaster. In RNA-polymerase, tRNA and Ribosomes: Their Genetics and Evolution, S. Osawa, H. Ozeki, H. Uchida and T. Yura, eds. (University of Tokyo Press, Tokyo), in press.

DeFranco, D., Schmidt, O. and Söll, D. (1980). Two control regions for eukaryotic tRNA gene transcription. Proc. Nat. Acad. Sci. USA, 77, 3365-3368.

Donis-Keller, H., Maxam, A.M. and Gilbert, W. (1977). Mapping adenines, guanines and pyrimidines in RNA. Nuc. Acids Res. 4, 2527-2538.

Elder, R.T., Szabo, P. and Uhlenbeck, O.C. (1980). 4S RNA gene organization in Drosophila melanogaster. In Transfer RNA: Biological Aspects, D. Söll, J. Abelson and P. Schimmel, eds. (Cold Spring Harbor Laboratory, New York), pp. 317-323.

Engelke, D.R., Ng, S., Shastry, B.S. and Roeder, R.G. (1980). Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. Cell 19, 717-728.

Etcheverry, T., Colby, D. and Guthrie, C. (1979). A precursor to a minor species of yeast $tRNA^{Ser}$ contains an intervening sequence. Cell 18, 11-26.

Feldmann, H. (1977). A comparison of transcriptional linkage of tRNA citrons in yeast and E. coli by the ultraviolet light technique. Nuc. Acids. Res. 4, 2831-2841.

Garber, R.L. and Gage, L.P. (1979). Transcription of a cloned Bombyx Mori $tRNA_2^{Ala}$ gene: Nucleotide sequence of the tRNA precursor and its processing in vitro. Cell 18, 817-828.

Goodman, H.M., Olson, M.V. and Hall, B.D. (1977). Nucleotide sequence of a mutant eukaryotic gene: The yeast tyrosine-inserting ochre suppressor SUP4-0. Proc. Nat. Acad. Sci. USA 74, 5453-5457.

Grunstein, M. and Hogness, D. (1975). Colony hybridization: A method for the isolation of a cloned DNAs that contain a specific gene. Proc. Nat. Acad. Sci. USA 72, 3961-3965.

Hayashi, S., Gillam, I.C., Delaney, A.D., Dunn, R., Tener, G.M., Grigliatti, T.A. and Suzuki, D.T. (1980). Hybridization of tRNAs of Drosophila melanogaster to polytene chromosomes. Chromosoma (Berl.) 76, 65-84.

Hershey, N.D., Conrad, S.E., Sodja, A., Yen, P.H., Cohen, Jr., M. Davidson, N., Ilgen, C. and Carbon, J. (1977). The sequence arrangement of Drosophila melanogaster 5S DNA cloned in recombinant plasmids. Cell 11, 585-598.

Hopper, A.K., Schultz, L.D. and Shapiro, R.A. (1980). Processing of intervening sequences: A new yeast mutant which fails to excise intervening sequences from precursor tRNAs. Cell 19, 741-751.

Hovemann, B., Sharp, S., Yamada, H. and Söll, D. (1980). Analysis of a Drosophila tRNA gene cluster. Cell 19, 889-895.

Johnson, J.D., Ogden, R., Johnson, P., Abelson, J., Dembeck, P. and Itakura, K. (1980). Transcription and processing of a yeast tRNA gene containing a modified intervening sequence. Proc. Nat. Acad. Sci. USA 77, 2564-2568.

Kang, H.S., Ogden, R.C. and Abelson, J. (1980). In 12th Miami Winter Symposium, cited by Johnson, et al. (1980) (Academic Press, New York), in press.

Knapp, G., Beckmann, J.S., Johnson, P.F., Fuhrman, S.A. and Abelson, J. (1978). Transcription and processing of intervening sequences in yeast tRNA genes. Cell 14, 221-236.

Knapp, G., Ogden, R.C., Peebles, C.L. and Abelson, J. (1979). Splicing of yeast tRNA precursors: Structure of the reaction intermediates. Cell 18, 37-45.

Korn, L.J. and Brown, D.D. (1978). Nucleotide sequence of Xenopus borealis oocyte 5S DNA: Comparison of sequences that flank several related eukaryotic genes. Cell 15, 1145-1156.

Kressman, A., Clarkson, S.G., Telford, J.L. and Birnstiel, M.L. (1977). Transcription of Xenopus $tRNA_1^{Met}$ and sea urchin histone DNA injected into the Xenopus oocyte nucleus. Cold Spring Harbor Symp. Quant. Biol. 42, 1077-1082.

Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975). Nucleotide sequence of the rightward operator of phage λ. Proc. Nat. Acad. Sci. USA 72, 1184-1188.

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C. Quon, D., Sim, G.K., and Efstradiatis, A. (1978). The isolation

of structural genes from libraries of eucaryotic DNA. Cell 15, 687-701.

Maxam, A. and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. In Methods in Enzymology, Vol. 65, L. Grossman and K. Moldave, eds. (Academic Press, New York), pp. 499-560.

Melton, D.A., DeRobertis, E.M. and Cortese, R. (1980). Order and intracellular location of the events involved in the maturation of a spliced tRNA. Nature 284, 143-148 (1980).

Mertz, J.E. and Davis, R.W. (1972). Cleavage of DNA by R_I restriction endonuclease generates cohesive ends. Proc. Nat. Acad. Sci. USA 69, 3370-3374.

Müller, F. and Clarkson, S.G. (1980). Nucleotide sequence of genes coding for $tRNA^{Phe}$ and $tRNA^{Tyr}$ from a repeating unit of X. laevis DNA. Cell 19, 345-353.

O'Farrell, P.Z., Cordell, B., Valenzuela, P., Rutter, W.J. and Goodman, H.M. (1978). Structure and processing of yeast precursor tRNAs containing intervening sequences. Nature 274 438-445.

Ogden, R.C., Beckmann, J.S., Abelson, J., Kang, H.S., Söll, D. and Schmidt, O. (1979). In vitro transcription and processing of yeast precursor tRNAs containing intervening sequences. Cell 17, 399-406.

Olson, M.V., Hall, B.D., Cameron, J.R. and Davis, R.W. (1979). Cloning of yeast tyrosine transfer RNA genes in bacteriophage lambda. J. Mol. Biol. 127, 285-295 (1979).

Peebles, C.L., Ogden, R.C., Knapp, G. and Abelson, J. (1979). Splicing of yeast tRNA precursors: A two-stage reaction. Cell 18, 27-35.

Radloff, R., Bauer, W. and Vinograd, J. (1967). A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA: The closed circular DNA in HeLa cells. Proc. Nat. Acad. Sci. USA 57, 1514-1521.

Roeder, R.G. (1976). Eukaryotic nuclear RNA polymerases. In RNA Polymerase, R. Losick and M. Chamberlain, eds. (Cold Spring Harbor Laboratory, New York), pp. 285-329.

Sakonju, S., Bogenhagen, D.F. and Brown, D.D. (1980). A control region in the center of the 5S RNA gene directs specific initiation of transcription: I. The 5' border of the region. Cell 19, 13-25.

Schmidt, O., Mao, J., Silverman, S., Hovemann, B. and Söll D. (1978). Specific transcription of eukaryotic tRNA genes in Xenopus germinal vesicle extracts. Proc. Nat. Acad. Sci. USA 75, 4819-4823.

Silverman, S., Schmidt, O., Söll, D. and Hovemann, B. (1979). The nucleotide sequence of a cloned Drosophila arginine tRNA gene and its in vitro transcription in Xenopus germinal vesicle extracts. J. Biol. Chem. 254, 10290-10294.

Smith, G.P. (1973). Unequal crossover and the evolution of multigene families. Cold Spring Harbor Symp. Quant. Biol. 38, 507-513.

Smith, G.P. (1976). Evolution of repeated DNA sequences by unequal crossover. Science 191, 528-535.

Smith, H.O. and Birnstiel, M.L. (1976). A simple method for DNA restriction site mapping. Nuc. Acids Res. 3, 2387-2398.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98 503-517.

Sprinzl, M., Grueter, F. Spelzhaus, A. and Gauss, D.H. (1980). Compilation of tRNA sequences. Nuc. Acids Res. 8, 1-22.

Staden, R. (1977). Sequence data handling by computer. Nuc. Acids Res. 4, 4037-4051.

Staden, R. (1978). Further procedures for sequence analysis by computer. Nuc. Acids Res. 5, 1013-1015.

Steffensen, D.M. and Wimber, D.E. (1971). Localization of tRNA genes in the salivary chromosomes of Drosophila by RNA:DNA hybridization. Genetics 69, 163-178.

Tener, G.M., Hayashi, S., Dunn, R., Delaney, A., Gillam, I.C., Grigliatti, T., Kaufman, T. and Suzuki, D.T. (1980). Transfer RNA genes of Drosophila melanogaster. In Transfer RNA: Biological Aspects, D. Söll, J. Abelson and P. Schimmel, eds. (Cold Spring Harbor Laboratory, New York), pp. 295-307.

Tinoco, Jr., I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralle, J. (1973). Improved estimation of secondary structure in ribonucleic acids. Nature New Biol. 246, 40-41.

Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R. and Rutter, W.J. (1978). Structure of yeast Phenylalanine-tRNA genes: An intervening DNA segment within the region coding for the tRNA. Proc. Nat. Acad. Sci. USA 75, 190-194.

Venegas, A., Quiroga, M., Zaldivar, J., Rutter, W.J. and Valenzuela, P. (1979). Isolation of yeast $tRNA^{Leu}$ genes: DNA sequence of a cloned $tRNA_3^{Leu}$ gene. J. Biol. Chem. 254, 12306-12309.

Weber, L. and Berger, E. (1976). Base sequence complexity of the stable RNA species of Drosophila melanogaster. Biochem. 15, 5511-5519.

Weil, P.A., Segall, J., Harris, B., Ng, S. and Roeder, R.G. (1979). Faithful transcription of eukaryotic genes by RNA polymerase III in systems reconstituted with purified DNA templates. J. Biol. Chem. 254, 6163-6173

Yen, P. and Davidson, N. (1980). The gross anatomy of a tRNA gene cluster at region 42A of the Drosophila melanogaster chromosome. Manuscript submitted.

Yen, P., Hershey, N.D., Robinson, R.R. and Davidson N. (1979). Sequence organization of Drosophila tRNA genes. In ICN-UCLA Symposium on Eucaryotic Gene Regulation, R. Axel and T. Maniatis, eds. (Academic Press, New York), pp. 133-142.

Yen, P.H., Sodja, A., Cohen, Jr., M., Conrad, S.E., Wu, M. Davidson, N. and Ilgen, C. (1977). Sequence arrangement of tRNA genes on a fragment of Drosophila melanogaster DNA cloned in E. coli. Cell 11, 763-777.

	T1 oligonucleotides from $tRNA^{Leu-a}$					T1 oligonucleotides from $tRNA^{Leu-b}$			
	predicted		observed			predicted		observed	
	^{32}P-GTP	^{32}P-UTP	^{32}P-GTP	^{32}P-UTP		^{32}P-GPT	^{32}P-UTP	^{32}P-GTP	^{32}P-UTP
a	20 (1)	20 (10)	22	22	a'	16 (1)	16 (7)	18	18
b	13 (1)	13 (4)	15	15	b	13 (1)	13 (4)	15	15
c	--	10-13 (4-7)	--	12,13	c'	--	11-14 (4-7)	--	13,14
d	9 (1)	9 (3)	10	10	d	9 (1)	9 (3)	10	10
e	7 (1)	7 (1)	7	7	e	7 (2)	7 (4)	7	7
f	6 (2)	6 (1)	6	6	f	6 (2)	6 (1)	6	6
g	5 (2)	5 (2)	5	5	g	5 (2)	5 (2)	5	5
h	4 (7)	4 (3)	4.3	4.3	h	4 (8)	4 (7)	4.3	4.3

Table 1. Predicted and observed lengths of RNAase T1 oligonucleotides from in vitro transcripts of the $tRNA^{Leu}$ genes. RNA lengths were determined from single strand DNA markers (Figure 6). Numbers in parenthesis indicate relative amount of radioactivity predicted in a band. Bands smaller than three nucleotides are not included. Bands c and c' are diagnostic bands, labeled by UTP but not GTP, from the 3' termini of transcription. These termini are assumed to be strings of U nucleotides immediately 3' to the tRNA coding sequences (see Discussion). Oligonucleotides originating from sequences 5' of the tRNA coding sequences are not included because the 5' termini of transcription have not been mapped.

tRNA gene	conserved sequence	# nucleotides matching Dm sequence	reference
X. laevis $tRNA^{Tyr}$	5' CAAUCCUU 3'	5	Müller and Clarkson, 1980
S. cerevisiae $tRNA^{Tyr}$	CAAAUCUU	7	Goodman et al., 1977
S. cerevisiae $tRNA^{Phe}$ (a)	AAAAACUU	7	Valenzuela et al., 1978
S. cerevisiae $tRNA^{Phe}$ (b)	AAAUACUU	6	Valenzuela et al., 1978
S. cerevisiae $tRNA^{Ser}_{UCG}$	AAAAAGUU	6	Etcheverry et al., 1979
S. cerevisiae $tRNA^{Trp}$	UAAAUCUU	7	Ogden et al., 1979
S. cerevisiae $tRNA^{Leu}_{3}$ (a)	AAUAUCUU	7	Kang et al., 1980
S. cerevisiae $tRNA^{Leu}_{3}$ (b)	AAAAUCUU	8	Venegas et al., 1979
D. melanogaster t RNA^{Leu}	AAAAUCUU		this report

Table 2. A conserved family of sequences in intervening sequences of tRNA genes. Underlined nucleotides correspond to RNA nucleotides predicted to be base paired in the tRNA precursor; overlined nucleotides correspond to RNA nucleotides predicted to be base paired to anticodon nucleotides in the tRNA precursor.

Figure Legends

Figure 1. Restriction and Southern hybridization analysis of Ch4:Dm50AB. Panel A - ethidium bromide stained gel of Ch4:Dm50AB digested with the indicated restriction enzymes (left); light and heavy exposures of Southern hybridization of the DNA in this gel with Dm^{32}P-tRNA probe (middle and right). The Eco Rl digest lane indicates that a single 10 kb fragment hybridizes to tRNA, while the Eco Rl + Hind III lane shows that two fragments of length 2.95 and 1.15 kb hybridize to tRNA. The 5.4 kb fragment in the Eco RI + Hind III lane that hybridizes to tRNA is due to incomplete Hind III digestion. Panel B - ethidium bromide stained gels of Ch4:Dm50AB digested with the indicated restriction enzymes and light and heavy exposures of Southern hybridizations using Dm^{32}P-tRNA probe. The Sal I + Bam HI lane (left) contains a single 4.3 kb fragment which hybridizes to tRNA (higher molecular weight fragments that hybridize to tRNA are due to incomplete Sal I digestion). This 4.3 kb fragment was subcloned into pBR322 to form p50AB. The Pvu II + Hind III lane (right) indicates that four distinct fragments hybridize to tRNA:1.75 kb; 1.11 kb; 0.66 kb; and 0.15 kb.

Figure 2. Restriction maps and sequencing gels of Ch4:Dm50AB and p50AB. Restriction enzyme code: A-Alu I; B-Bam HI; D-Hind III; E-Hae III; F-Hinf I; H-Hha I; HP-HpaI; M-MspI; P-Pvu II; Rl-Eco RI; S-Sal I; T-Taq I; V-AvaII. Sequencing gels are schematically shown

as arrows in the direction of reading; dashed arrows indicate 3' end labeled fragments, while solid arrows indicate 5' end labeled fragments. tRNAs are schematically shown as arrows indicating direction of transcription.

Figure 3. Sequences of p50AB tRNA genes and their flanking sequences. Panel A - The nontranscribed DNA strand in the 5' to 3' direction is shown. All five $tRNA^{Ile}$ genes have identical sequences for mature $tRNA^{Ile}$. The $tRNA^{Leu}$ genes differ by a T→G transversion at position 42, and a 7 bp deletion starting at position 70. ▲ indicates proposed sites for excision of the intervening sequences. Underlined sequences indicate small, strongly conserved sequences preceding the $tRNA^{Ile}$ and $tRNA^{Leu}$ genes. Panel B - Clover leaf structures for p50AB $tRNA^{Ile}$ and $tRNA^{Leu}$. The $tRNA^{Leu}$ structures contain the intervening sequence nucleotides. For the proposed stable secondary structure of the intervening sequences of the $tRNA^{Leu}$ genes, see Figure 7.

Figure 4. In vitro transcription products of p50AB tRNA genes. In vitro transcription products were run on denaturing 8% acrylamide gels and exposed to X-ray film. Panel A - lanes a, b, and c contain the products of in vitro transcription of supercoiled plasmid DNAs with the indicated ^{32}P label: lane a - p50AB DNA (α-^{32}P-UTP); lane b - pBR322 DNA (α-^{32}P-GTP); lane c - pBR322 DNA (α-^{32}P-UTP). Lane d contains ^{32}P labeled Drosophila tRNA. Panel B - restriction fragments from p50AB DNA containing the

indicated tRNA genes were transcribed in vitro with α-^{32}-P-GTP label: lane a - 588 bp Alu I fragment ($tRNA^{Leu-a}$); lane b - 644 bp Alu I fragment ($tRNA^{Leu-b}$); lane c - 674 bp Pvu II fragment ($tRNA^{Ile-d}$); lane d - 169 bp Pvu II + Hind III fragment ($tRNA^{Ile-e}$); lane e - 618 bp Hha I fragment ($tRNA^{Ile-a}$); lane f - 1773 bp Hha I fragment ($tRNA^{Ile-b,c,d,e}$).

Figure 5. Chase of 137 and 145 base transcripts with in vitro transcription extract. The 137 base and 145 base transcripts (lanes a and b of Figure 4) were excised and eluted. The RNAs were treated with an in vitro transcription extract containing unlabeled UTP. At the indicated times, aliquots were removed and quenched as before. Lanes a-d (137 base transcript): lane a - 60 min; lane b - 10 min.; lane c - 1 min.; lane d - 0 min., lanes e-f (145 base transcript): lane e - 60 min.; lane f - 10 min.; lane g - 1 min.; lane h - 0 min.

Figure 6. RNAase T1 digestion of the 137 and 145 base transcripts. 137 and 145 base transcripts, which were transcribed from isolated restriction fragments with either α-^{32}P-GTP or α-^{32}P-UTP label, were digested with RNAase T1 and loaded onto an 8 M urea, 25% acrylamide gel with single stranded DNA length standards: lane a - 137 base transcript (α-^{32}P-UTP); lane b - 137 base transcript (α-^{32}P-GTP); lane c - 145 base transcript (α-^{32}P-UTP); lane d - 145 base transcript (α ^{32}P-GTP). See Table 1 for analysis of this gel. The numbers at left are nucleotide positions

from a Maxam and Gilbert (1980) sequencing gel reading. Because of the chemical degradation method of sequencing, however, nucleotide position n is actually an oligonucleotide of length n - 1. Thus, the RNA oligonucleotide migrating at position 23 comigrates with DNA oligonucleotides of length 22.

Figure 7. Predicted secondary structures of the precursor of a yeast $tRNA_3^{Leu}$ and the proposed precursors of Drosophila $tRNA^{Leu}$. Underlines indicate short regions of strong sequence homology between the intervening sequences of yeast and Drosophila $tRNA^{Leu}$. Nucleotides in parentheses indicate differences between the sequences for p50AB $tRNA^{Leu-b}$ and $tRNA^{Leu-a}$. Heavy arrows indicate predicted locations of splice junctions, and the predicted mature anticodon stem and loop is shown to the right of each precursor.

Figure 8. Comparison of the sequences of yeast $tRNA_3^{Leu}$ and p50AB $tRNA^{Leu-b}$. Loopouts were allowed in this comparison to maximize homology (looped out nucleotides are opposite -). Matches between the two sequences are indicated by *.

Figure 9. Homologies in the DNA sequences 5' to Drosophila $tRNA^{Ile}$ genes. 50 bases preceding each Dm $tRNA^{Ile}$ gene reported here ($tRNA^{Ile\text{-a to e}}$) and by Hovemann, et al. (1980) ($tRNA^{Ile\text{-bh}}$) were compared for homologies. Alignments producing maximum homology were found by a computer program that does not allow

loopouts in comparisons. Each alignment is identified by the tRNA genes that follow the compared DNA sequences: a - p50AB $tRNA^{Ile-a}$; b - p50AB $tRNA^{Ile-b}$; c - p50AB $tRNA^{Ile-c}$; d - p50AB $tRNA^{Ile-d}$; e - p50AB $tRNA^{Ile-e}$; bh - pCIT12 $tRNA^{Ile}$.

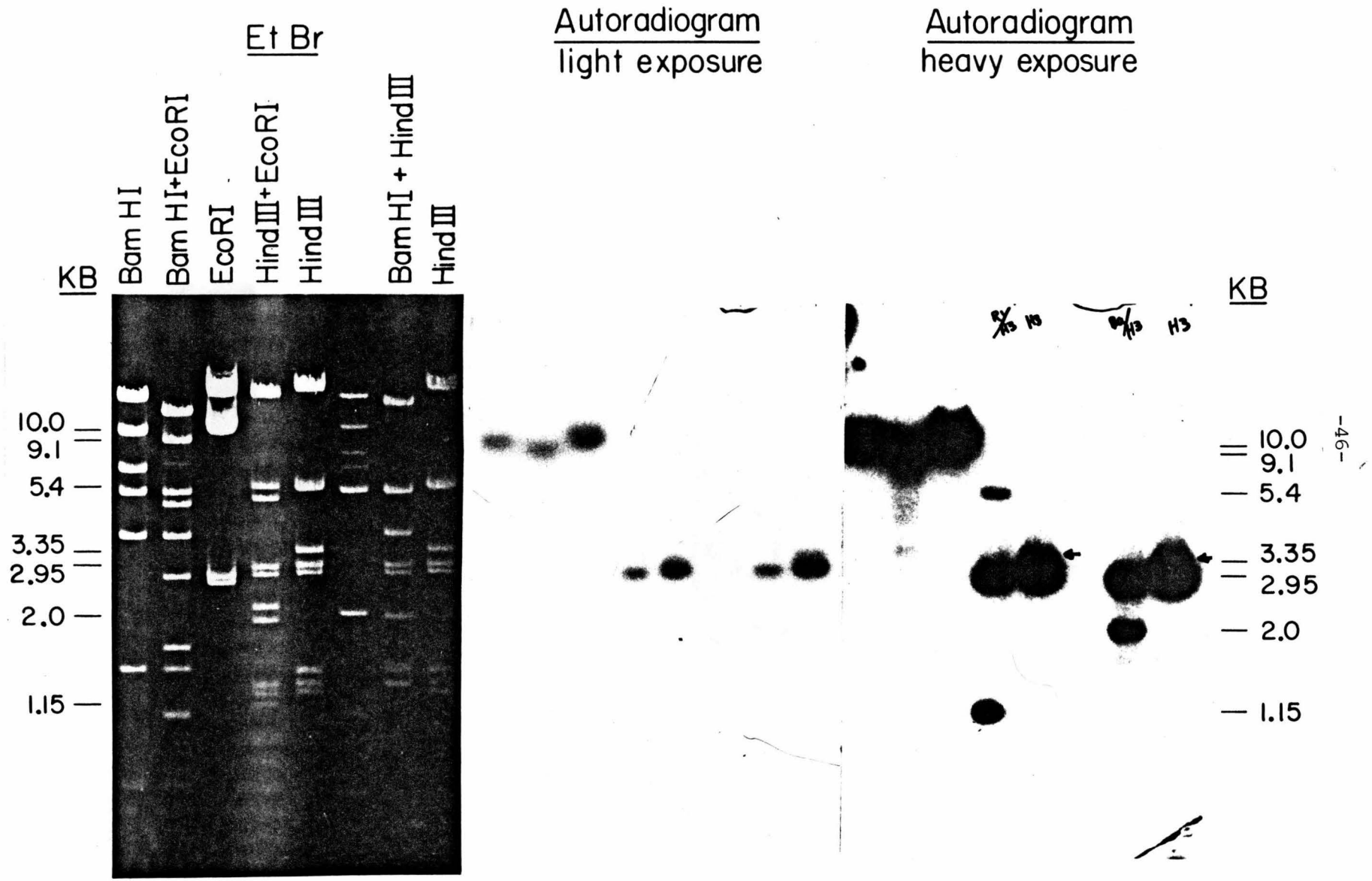

Et Br
Autoradiogram
light exposure
Autoradiogram
heavy exposure
KB
Bam HI
Bam HI+EcoRI
EcoRI
HindIII+EcoRI
HindIII
Bam HI + HindIII
HindIII
10.0
9.1
5.4
3.35
2.95
2.0
1.15
KB
10.0
9.1
5.4
3.35
2.95
2.0
1.15

Figure 1A

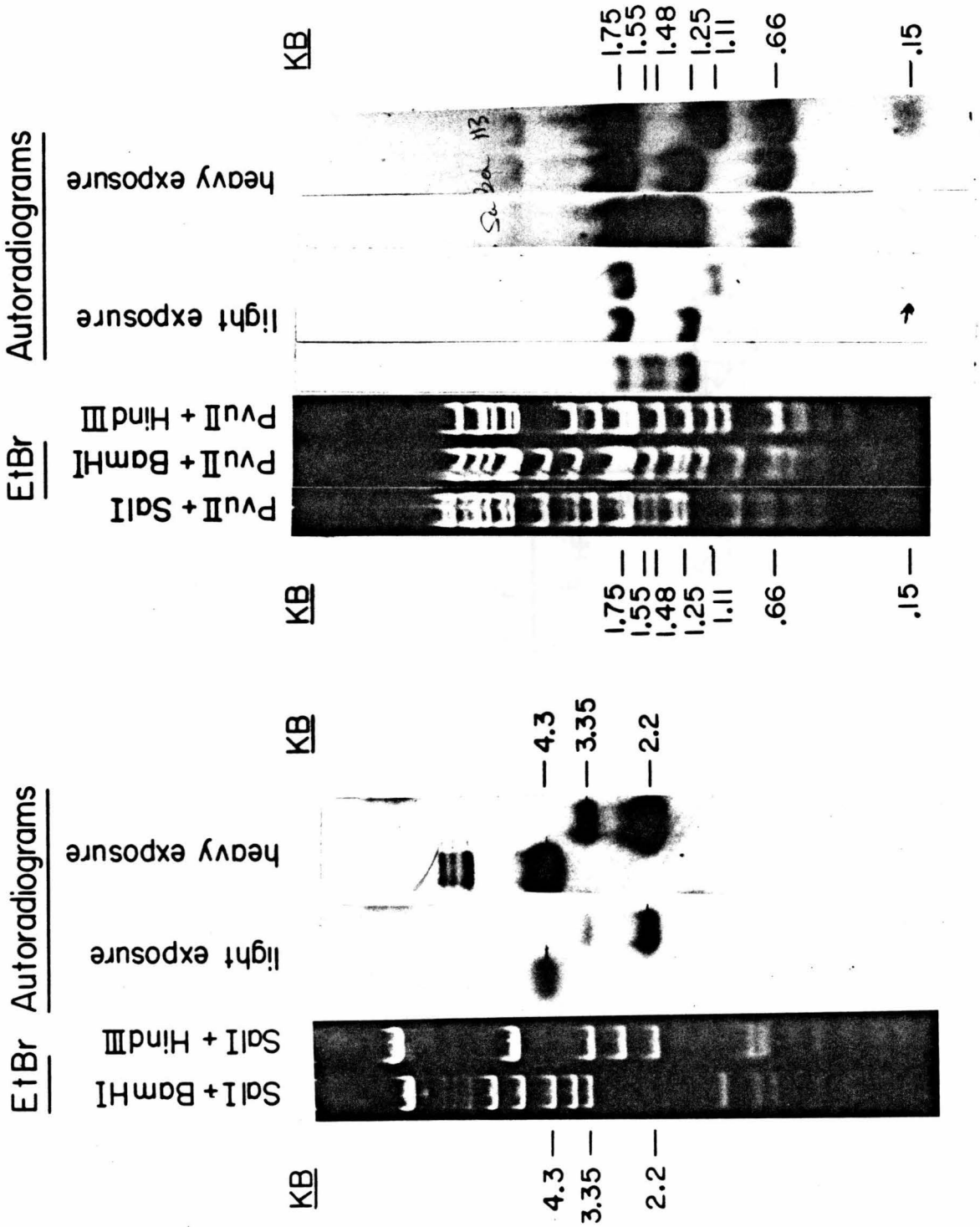

EtBr
SalI + BamHI
SalI + HindIII
Autoradiograms
light exposure
heavy exposure
KB
4.3
3.35
2.2
EtBr
PvuII + SalI
PvuII + BamHI
PvuII + HindIII
Autoradiograms
light exposure
heavy exposure
KB
1.75
1.55
1.48
1.25
1.11
.66
.15

Figure 1B

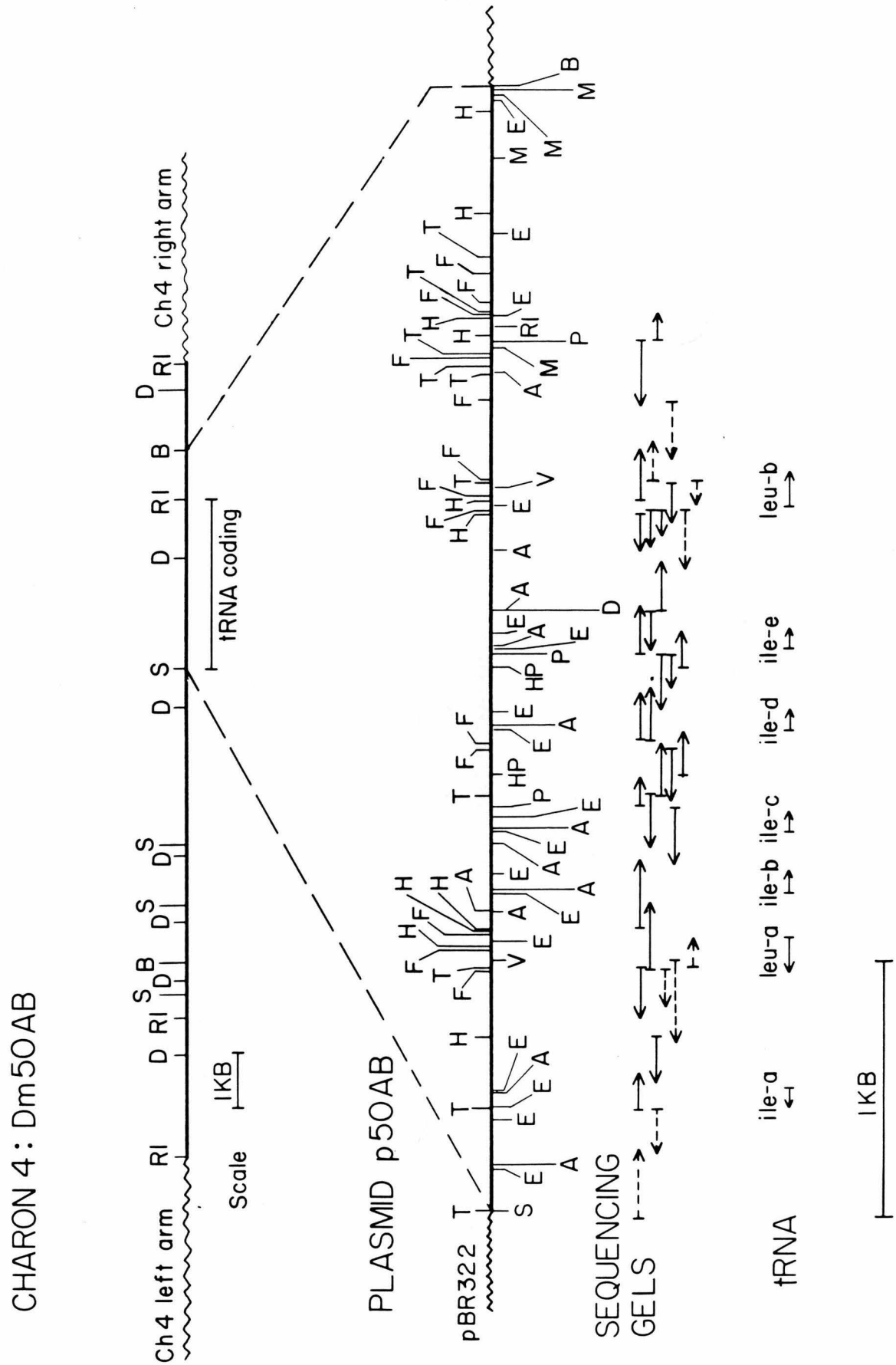

CHARON 4 : Dm50AB
Ch4 left arm
Ch4 right arm
tRNA coding
Scale
1KB
PLASMID p50AB
pBR322
SEQUENCING GELS
tRNA
ile-a
leu-a
ile-b
ile-c
ile-d
ile-e
leu-b
1KB

Figure 2

5' FLANKING SEQUENCES

```
     -100                                                -50                                                -1
[ILE-A] TGCCCACTTAAAATTTATTAGGCAAAGTTTTTTGTTTTGCTATGATTTTCAGTTTAGTACGCCTTCCAGTCTGCTACGCCTCTTGCATCCGTAAATACAT
[ILE-B] AATTCTATTCGTAATACAACAGTTGCAGTCAAACCAACATGGAAGGTATTGTTAGCTTCGCTATTTAGCGAAAATCGCATTTTGTTAAAATTACAGACAT
[ILE-C] TTTACGGATAGAAATTTATGTCTTTTACACAACTTGTGGATAAAAGCTCAATATTGTGTTTCATTGGCGTAAATGCGTCTTTTGACAGAAATCAGCACAA
[ILE-D] CGTTGAGGTCCTATGCATTCCGAATCTTTGGTTCAGTGACCAATTCTACGGCTGTTCGTCGCATTTACTCTCAAAAAGCGTTTGGCAAAAAAGAACGCAT
[ILE-E] TACGCTCCCAACATAATATACTACTGTTAAATACAACTGTTAACAGATTTATCAGAATTGCTATTTGGTAAAACTGGCATTTTGGCAGCTGTGCATAAAT
[LEU-A] TGGTTTGACTGCAACTGTTGTATTACGAATAGAATTGTTCTAGTACTGGCACATGCGCTGGCTCGCGCGCTACTACGGAATCGGACACGAGTTTGGCTAT
[LEU-B] TAGACAATATAATGGCAAACAGAGAACATCCTAATTTATTGGTTCAACGCCCCATTTTGTATATACTTTTATGACACTAGATTTTGGGCGCGGAATCAGT
```

TRNA SEQUENCES (5'-3')

```
          1                   20                  40                  60                  80                  100                 120
[ILE]     GGCCCATTAGCTCAGTTGGTTAGAGCGTCGTGCTAATAACGCGAAGGTCGCGGGTTCGATCCCCTCATGGGCCA
[LEU-A]   GTCAGGATGGCCGAGCGGTCTAAGGCGCCAGACTCAAGATTTAAAATCTTACTTTCTGAACGAAAGTGT ... ATGAGCGTTCTGGTCCTCTCTGAGGGCGTGGGTTCGAATCCCACTTCTGACA
[LEU-B]   GTCAGGATGGCCGAGCGGTCTAAGGCGCCAGACTCAAGATTGAAAATCTTACTTTCTGAACGAAAGTGTTTGTTTAATGAGCGTTCTGGTCCTCTCTGAGGGCGTGGGTTCGAATCCCACTTCTGACA
```

3' FLANKING SEQUENCES

```
          +1                                                  +50
[ILE-A] GTCGAACTTTTTTTATAATGCTTTACGTTTTTTAGTAGATGCCTAAGGAT
[ILE-B] GGTGAATTTTTTGTATCATCCCCCCCACAAGGTCTTTTAAAGATACTAAG
[ILE-C] AACATTCTTTTTTCTTGTTATATACATATTTACAGCTGAATAATATACAT
[ILE-D] GATATTTTTTATTTCTGAACATCCTGATTACTTTGAGGTTCAAAGCCTTG
[ILE-E] GAATAATTTTTTTTTATTACATTTCTAATGCACTTTGAAAATTAAAAAGT
[LEU-A] AAATTTTTTTAGTGGTTTTGTTTTTGAATGGTTGATAAATTTTTTCTACC
[LEU-B] AATATTTTTTTAAGCCCAAAGAAATTAACAATGATTTGACAAACAATGCA
```

Figure 3A

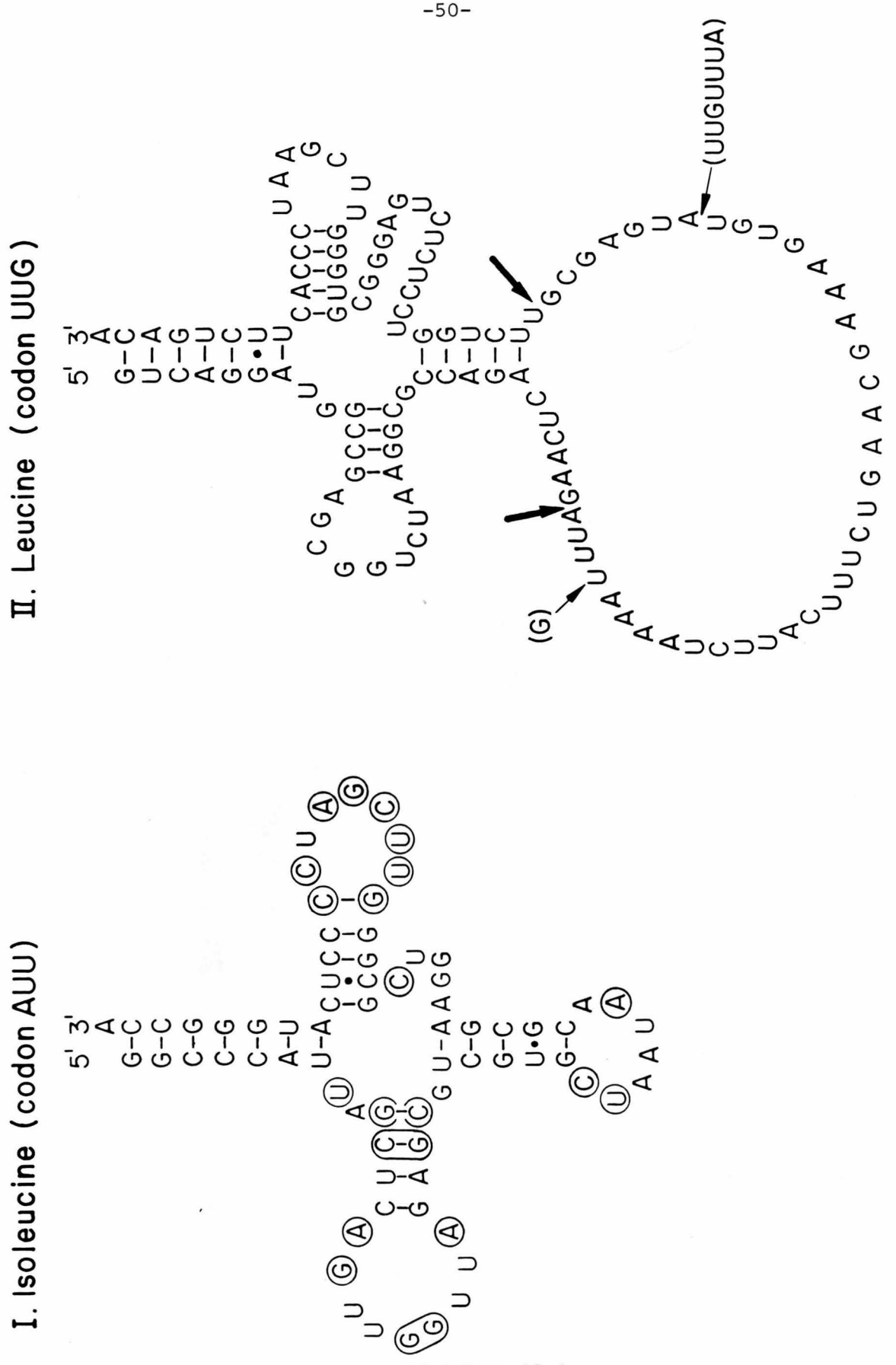


Figure 3B

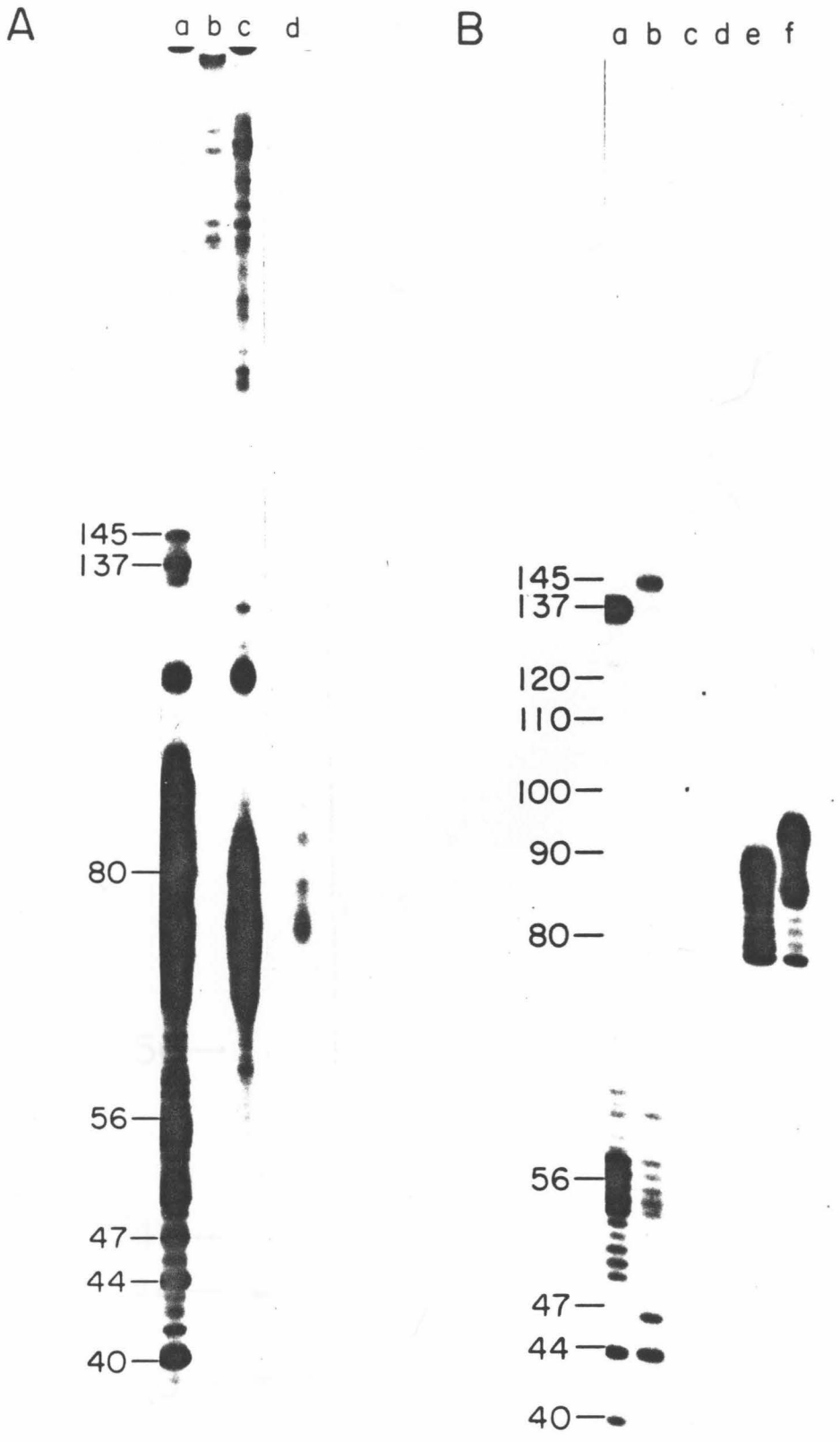

A
a b c d
145
137
80
56
47
44
40
B
a b c d e f
145
137
120
110
100
90
80
56
47
44
40

Figure 4

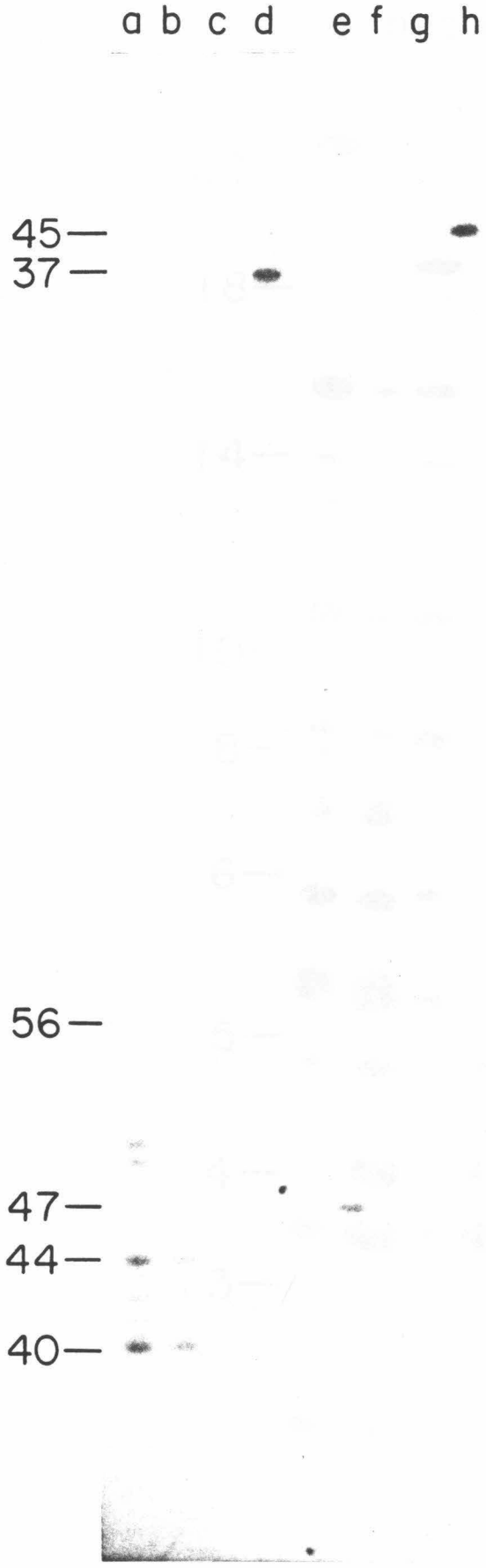


Figure 5

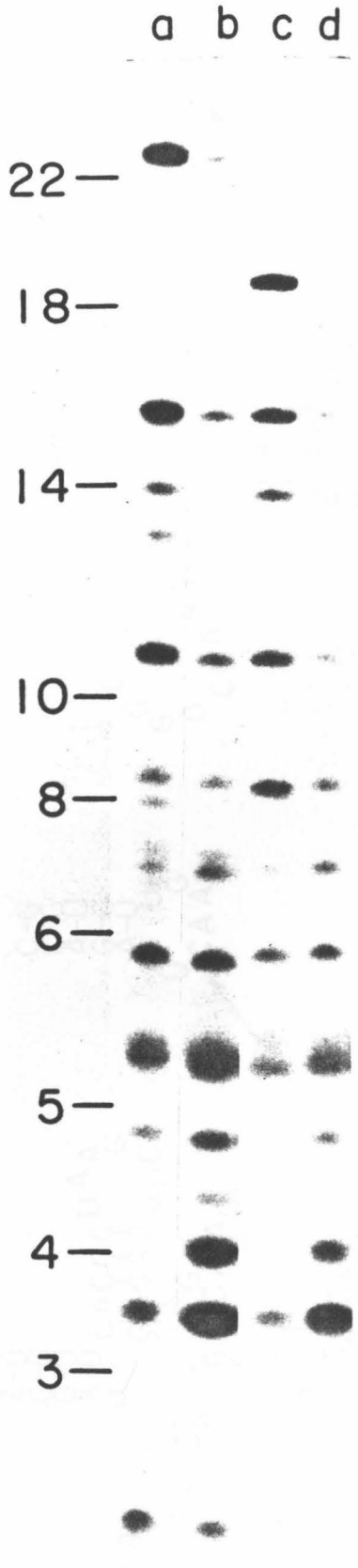
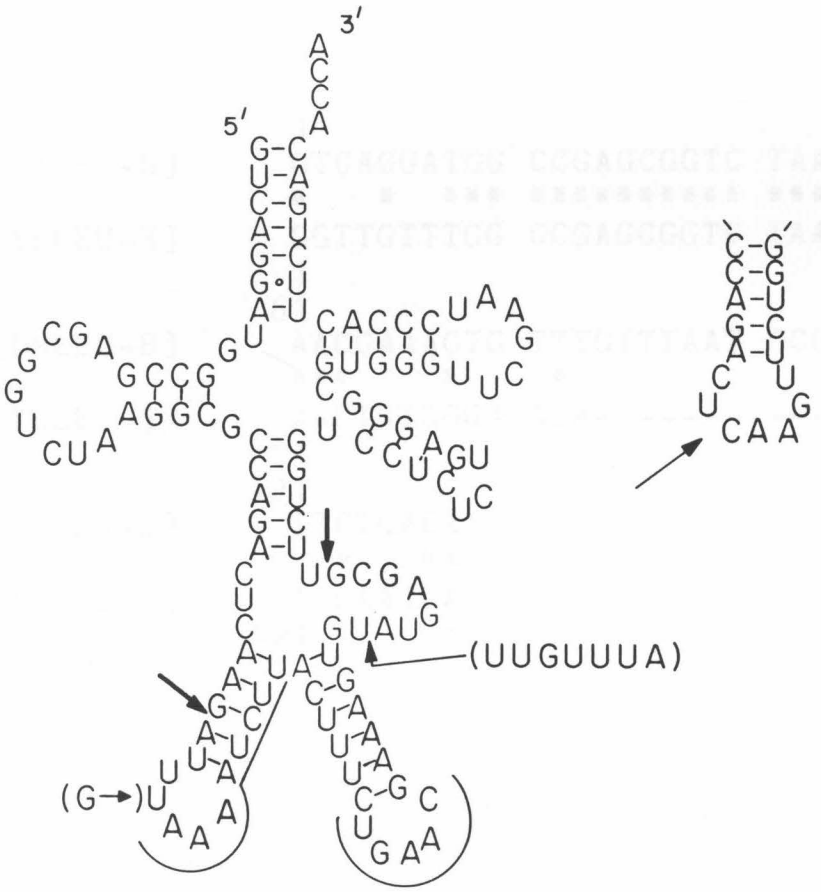


Figure 6

Proposed precursors to Drosophila $tRNA^{Leu}$



Precursor to yeast $tRNA^{Leu}_3$ (Venegas, et al., 1979)

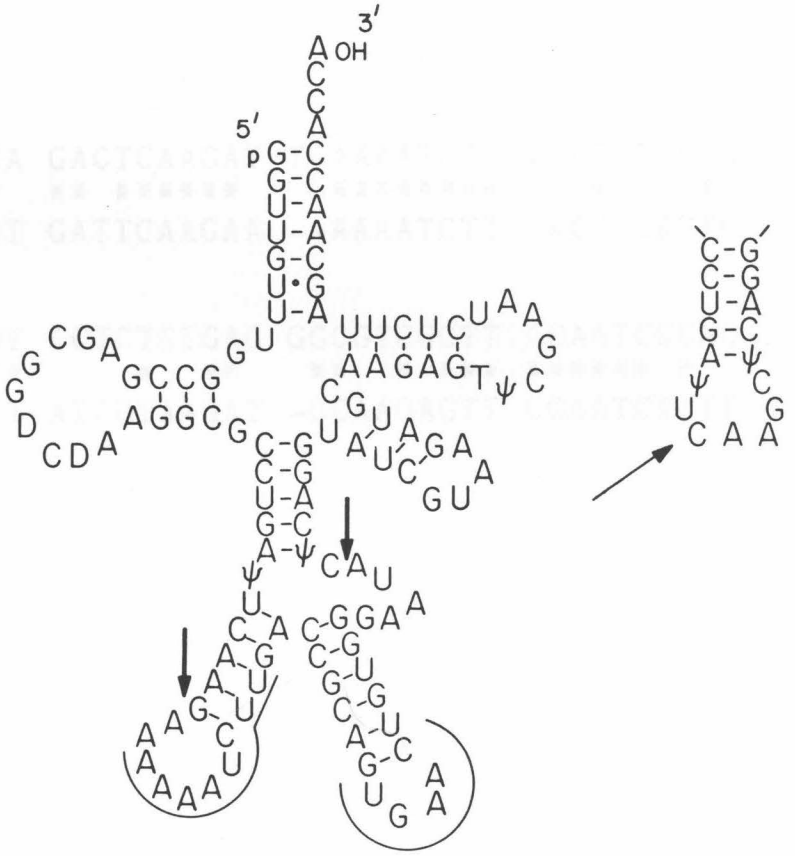


Figure 7

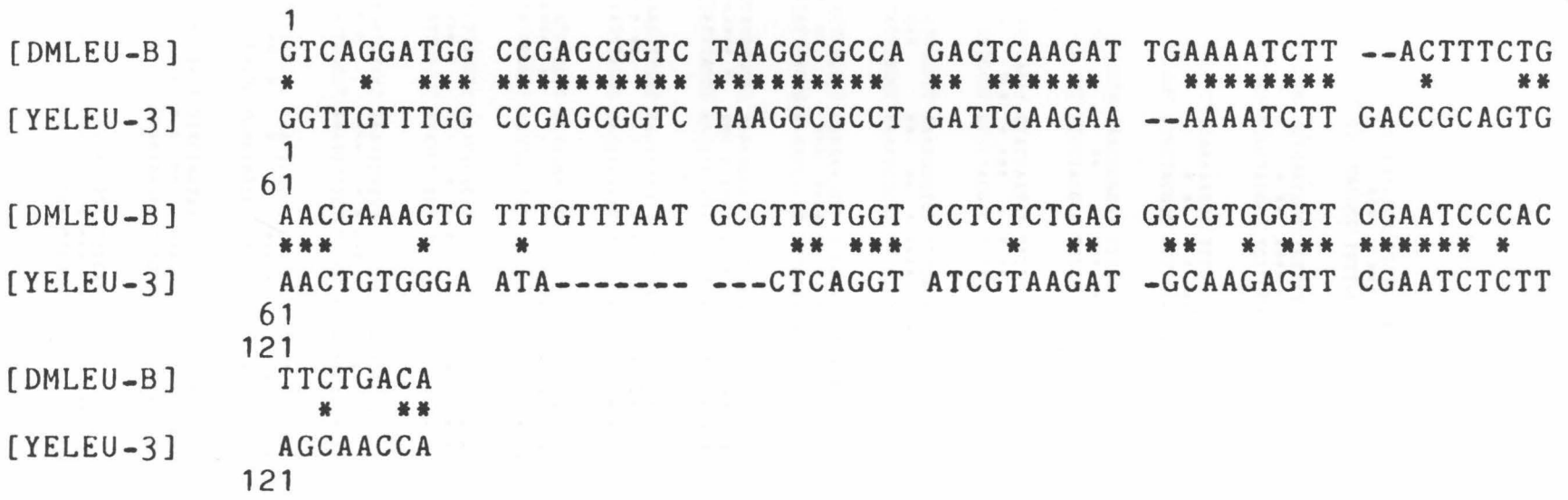

1
[DMLEU-B] GTCAGGATGG CCGAGCGGTC TAAGGCGCCA GACTCAAGAT TGAAAATCTT --ACTTTCTG
[YELEU-3] GGTTGTTTGG CCGAGCGGTC TAAGGCGCCT GATTCAAGAA --AAAATCTT GACCGCAGTG
1
61
[DMLEU-B] AACGAAAGTG TTTGTTTAAT GCGTTCTGGT CCTCTCTGAG GGCGTGGGTT CGAATCCCAC
[YELEU-3] AACTGTGGGA ATA------- ---CTCAGGT ATCGTAAGAT -GCAAGAGTT CGAATCTCTT
61
121
[DMLEU-B] TTCTGACA
[YELEU-3] AGCAACCA
121

Figure 8

	5′	3′	tRNA^{Ile} gene	% homology
1.	GTTAGCTTC GCTATTTAGC GAAAATCGCA TTTTGTTAAA ATTACAGACAT		b	38
	AGTTTAGTAC GCCTTCCAGT CTGCTACGCC TCTTGCATCC GTAAATACAT		a	
2.	ATATTGT GTTTCATTGG CGTAAATGCG TCTTTTGACA GAAATCAGCACAA		c	38
	AGTTTAGTAC GCCTTCCAGT CTGCTACGCC TCTTGCATCC GTAAATACAT		a	
3.	GCTGTTCGTC GCATTTACTC TCAAAAAGCG TTTGGCAAAA AAGAACGCAT		d	40
	AGTTTAGTAC GCCTTCCAGT CTGCTACGCC TCTTGCATCC GTAAATACAT		a	
4.	ATCAGAATTG CTATTTGGTA AAACTGGCAT TTTGGCAGCT GTGCATAAAT		e	40
	AGTTTAGTAC GCCTTCCAGT CTGCTACGCC TCTTGCATCC GTAAATACAT		a	
5.	TTCTAAGC TTCACAAATT TTACTAATAT TTTTTTTCAA TGCATTCCATGG		bh	36
	AGTTTAGTAC GCCTTCCAGT CTGCTACGCC TCTTGCATCC GTAAATACAT		a	
6.	ATATTGTGTT TCATTGGCGT AAATGCGTCT TTTGACAGAA ATCAGCACAA		c	44
	GTTAGCTTCG CTATTTAGCG AAAATCGCAT TTTGTTAAAA TTACAGACAT		b	
7.	GCTGTTCGTCG CATTTACTCT CAAAAAGCGT TTGGCAAAAA AGAACGCAT		d	48
	GTTAGCTTCG CTATTTAGCG AAAATCGCAT TTTGTTAAAA TTACAGACAT		b	
8.	ATCAGAATTG CTATTTGGTA AAACTGGCAT TTTGGCAGCT GTGCATAAAT		e	62
	GTTAGCTTCG CTATTTAGCG AAAATCGCAT TTTGTTAAAA TTACAGACAT		b	
9.	TTCTAAGCTTCA CAAATTTTAC TAATATTTTT TTTCAATGCA TTCCATGG		bh	44
	GTTAGCTTCG CTATTTAGCG AAAATCGCAT TTTGTTAAAA TTACAGACAT		b	
10.	GCTGTTCGTC GCATTTACTC TCAAAAAGCG TTTGGCAAAA AAGAACGCAT		d	46
	ATATTGTGTT TCATTGGCGT AAATGCGTCT TTTGACAGAA ATCAGCACAA		c	
11.	ATCAGAATTG CTATTTGGTA AAACTGGCAT TTTGGCAGCT GTGCATAAAT		e	44
	ATATTGTGTT TCATTGGCGT AAATGCGTCT TTTGACAGAA ATCAGCACAA		c	
12.	TTCTAAGCTTCA CAAATTTTAC TAATATTTTT TTTCAATGCA TTCCATGG		bh	36
	ATATTGTGTT TCATTGGCGT AAATGCGTCT TTTGACAGAA ATCAGCACAA		c	
13.	ATCAGAATTG CTATTTGGTA AAACTGGCAT TTTGGCAGCT GTGCATAAAT		e	36
	GCTGTTCGTC GCATTTACTC TCAAAAAGCG TTTGGCAAAA AAGAACGCAT		d	
14.	TTCTAAGC TTCACAAATT TTACTAATAT TTTTTTTCAA TGCATTCCATGG		bh	36
	GCTGTTCGTC GCATTTACTC TCAAAAAGCG TTTGGCAAAA AAGAACGCAT		d	
15.	TTCTAAGCTTCA CAAATTTTAC TAATATTTTT TTTCAATGCA TTCCATGG		bh	40
	ATCAGAATTG CTATTTGGTA AAACTGGCAT TTTGGCAGCT GTGCATAAAT		e	

Figure 9

APPENDIX I

Nucleotide Sequence of Drosophila Sequences of p50AB: Sequence from the Sal I site extending 50 nucleotides beyond the Eco RI site.

```
        10         20         30         40         50         60         70         80         90        100        110        120
GTCGACGAGT GGGAAAGCAA CAGAGCAACT CAACATCAAG TGCTTGCCTT GGCGGGGTGG GGCGGGTGGA TTTTGATTTT TGTATTAGCG GGCGGCGGTT GGTTGCCAGA CCGCGGAAAG
CAGCTGCTCA CCCTTTCGTT GTCTCGTTGA GTTGTAGTTC ACGAACGGAA CCGCCCCACC CCGCCCACCT AAAACTAAAA ACATAATCGC CCGCCGCCAA CCAACGGTCT GGCGCCTTTC

       130        140        150        160        170        180        190        200        210        220        230        240
TAAGCGGCCA AGAAATCAAT TTAAGCTAGC TTTAGCCCTT TGTTAATGGG TTATCAAGAT TTACATGCAC GCAAAACGCT TGGTAATTAG TTATCACACA TCTGGATGTC AAAACTGGAA
ATTCGCCGGT TCTTTAGTTA AATTCGATCG AAATCGGGAA ACAATTACCC AATAGTTCTA AATGTACGTG CGTTTTGCGA ACCATTAATC AATAGTGTGT AGACCTACAG TTTTGACCTT

       250        260        270        280        290        300        310        320        330        340        350        360
GGCCAAGTGT CAGACTGACA AATGATATGA AAGGTCTAAG GACCAAGCCT AAGGGTATTC ATCCTTAGGC ATCTACTAAA AAACGTAAAG CATTATAAAA AAAGTTCGAC TGGCCCATGA
CCGGTTCACA GTCTGACTGT TTACTATACT TTCCAGATTC CTGGTTCGGA TTCCCATAAG TAGGAATCCG TAGATGATTT TTTGCATTTC GTAATATTTT TTTCAAGCTG ACCGGGTACT

       370        380        390        400        410        420        430        440        450        460        470        480
GGGGATCGAA CCCGCGACCT TCGCGTTATT AGCACGACGC TCTAACCAAC TGAGCTAATG GGCCATGTAT TTACGGATGC AAGAGGCGTA GCAGACTGGA AGGCGTACTA AACTGAAAAT
CCCCTAGCTT GGGCGCTGGA AGCGCAATAA TCGTGCTGCG AGATTGGTTG ACTCGATTAC CCGGTACATA AATGCCTACG TTCTCCGCAT CGTCTGACCT TCCGCATGAT TTGACTTTTA

       490        500        510        520        530        540        550        560        570        580        590        600
CATAGCAAAA CAAAAAACTT TGCCTAATAA ATTTTAAGTG GGCATTAATT TAAACAATTT AATATACAAT ACTTGTTAAG TTTTTGCGCG AAATTTGTGG AATAAAGTTA ATAAGTTAGT
GTATCGTTTT GTTTTTTGAA ACGGATTATT TAAAATTCAC CCGTAATTAA ATTTGTTAAA TTATATGTTA TGAACAATTC AAAAACGCGC TTTAAACACC TTATTTCAAT TATTCAATCA

       610        620        630        640        650        660        670        680        690        700        710        720
CATTTTCATA CTGTAATCCC CCTTAAAAAT AAATTTTATA ATTATGTTAT GATTGAACAA CCGCAATTCT GTATTTCTTA AATTTTATTG TAAAATATAT TTTTATAGGT TAGGTAGAAA
GTAAAAGTAT GACATTAGGG GGAATTTTTA TTTAAAATAT TAATACAATA CTAACTTGTT GGCGTTAAGA CATAAAGAAT TTAAAATAAC ATTTTATATA AAAATATCCA ATCCATCTTT

       730        740        750        760        770        780        790        800        810        820        830        840
AAATTTATCA ACCATTCAAA AACAAAACCA CTAAAAAAAT TTTGTCAGAA GTGGGATTCG AACCCACGCC CTCAGAGAGG ACCAGAACGC TCATACACTT TCGTTCAGAA AGTAAGATTT
TTTAAATAGT TGGTAAGTTT TTGTTTTGGT GATTTTTTTA AAACAGTCTT CACCCTAAGC TTGGGTGCGG GAGTCTCTCC TGGTCTTGCG AGTATGTGAA AGCAAGTCTT TCATTCTAAA

       850        860        870        880        890        900        910        920        930        940        950        960
TAAATCTTGA GTCTGGCGCC TTAGACCGCT CGGCCATCCT GACATAGCCA AACTCGTGTC CGATTCCGTA GTAGCGCGCG AGCCAGCGCA TGTGCCAGTA CTAGAACAAT TCTATTCGTA
ATTTAGAACT CAGACCGCGG AATCTGGCGA GCCGGTAGGA CTGTATCGGT TTGAGCACAG GCTAAGGCAT CATCGCGCGC TCGGTCGCGT ACACGGTCAT GATCTTGTTA AGATAAGCAT

       970        980        990       1000       1010       1020       1030       1040       1050       1060       1070       1080
ATACAACAGT TGCAGTCAAA CCAACATGGA AGGTATTGTT AGCTTCGCTA TTTAGCGAAA ATCGCATTTT GTTAAAATTA CAGACATGGC CCATTAGCTC AGTTGGTTAG AGCGTCGTGC
TATGTTGTCA ACGTCAGTTT GGTTGTACCT TCCATAACAA TCGAAGCGAT AAATCGCTTT TAGCGTAAAA CAATTTTAAT GTCTGTACCG GGTAATCGAG TCAACCAATC TCGCAGCACG
```

```
      1090        1100        1110        1120        1130        1140        1150        1160        1170        1180        1190        1200
TAATAACGCG AAGGTCGCGG GTTCGATCCC CTCATGGGCC AGGTGAATTT TTTGTATCAT CCCCCCCACA AGGTCTTTTA AAGATACTAA GGAAAATCAC CGAATTATAT GATAATTTTG
ATTATTGCGC TTCCAGCGCC CAAGCTAGGG GAGTACCCGG TCCACTTAAA AAACATAGTA GGGGGGGTGT TCCAGAAAAT TTCTATGATT CCTTTTAGTG GCTTAATATA CTATTAAAAC

      1210        1220        1230        1240        1250        1260        1270        1280        1290        1300        1310        1320
AAAAGAATAT GCTGAACGCA TCATTGAAAT CCTTTATGC- TTTTATTTCA TTTTACGGAT AGAAATTTAT GTCTTTTACA CAACTTGTGG ATAAAAGCTC AATATTGTGT TTCATTGGCG
TTTTCTTATA CGACTTGCGT AGTAACTTTA GGAAATACG- AAAATAAAGT AAAATGCCTA TCTTTAAATA CAGAAAATGT GTTGAACACC TATTTTCGAG TTATAACACA AAGTAACCGC

      1330        1340        1350        1360        1370        1380        1390        1400        1410        1420        1430        1440
TAAATGCGTC TTTTGACAGA AATCAGCACA AGGCCCATTA GCTCAGTTGG TTAGAGCGTC GTGCTAATAA CGCGAAGGTC GCGGGTTCGA TCCCCTCATG GGCCAAACAT TCTTTTTTCT
ATTTACGCAG AAAACTGTCT TTAGTCGTGT TCCGGGTAAT CGAGTCAACC AATCTCGCAG CACGATTATT GCGCTTCCAG CGCCCAAGCT AGGGGAGTAC CCGGTTTGTA AGAAAAAAGA

      1450        1460        1470        1480        1490        1500        1510        1520        1530        1540        1550        1560
TGTTATATAC ATATTTACAG CTGAATAATA TACATAATTA ATAAATAATA ATAATCCCTC GAAAATAGTT CTCAAGCAGT GATTTCAATA TGCGTACTTA TGAACTCAAT TATTGAGACA
ACAATATATG TATAAATGTC GACTTATTAT ATGTATTAAT TATTTATTAT TATTAGGGAG CTTTTATCAA GAGTTCGTCA CTAAAGTTAT ACGCATGAAT ACTTGAGTTA ATAACTCTGT

      1570        1580        1590        1600        1610        1620        1630        1640        1650        1660        1670        1680
AAATAGTTAA CTTGTACATA GGCACACCAT TTTTGTGAAT TTAACAAATG CTAAGAATAC CTAGGGTATT TAAGTCATCA CTAGTTCTTT TTTACCCTAG GAGAGACAAT TTACTGATCA
TTTATCAATT GAACATGTAT CCGTGTGGTA AAAACACTTA AATTGTTTAC GATTCTTATG GATCCCATAA ATTCAGTAGT GATCAAGAAA AAATGGGATC CTCTCTGTTA AATGACTAGT

      1690        1700        1710        1720        1730        1740        1750        1760        1770        1780        1790        1800
CGAATCCGTT TGCTTGGTAT TGACTAAGAA ATGGGAATTA TCAGGAATTG TCCGACGTTG AGGTCCTATG CATTCCGAAT CTTTGGTTCA GTGACCAATT CTACGGCTGT TCGTCGCATT
GCTTAGGCAA ACGAACCATA ACTGATTCTT TACCCTTAAT AGTCCTTAAC AGGCTGCAAC TCCAGGATAC GTAAGGCTTA GAAACCAAGT CACTGGTTAA GATGCCGACA AGCAGCGTAA

      1810        1820        1830        1840        1850        1860        1870        1880        1890        1900        1910        1920
TACTCTCAAA AAGCGTTTGG CAAAAAAGAA CGCATGGCCC ATTAGCTCAG TTGGTTAGAG CGTCGTGCTA ATAACGCGAA GGTCGCGGGT TCGATCCCCT CATGGGCCAG ATATTTTTTA
ATGAGAGTTT TTCGCAAACC GTTTTTTCTT GCGTACCGGG TAATCGAGTC AACCAATCTC GCAGCACGAT TATTGCGCTT CCAGCGCCCA AGCTAGGGGA GTACCCGGTC TATAAAAAAT

      1930        1940        1950        1960        1970        1980        1990        2000        2010        2020        2030        2040
TTTCTGAACA TCCTGATTAC TTTGAGGTTC AAAGCCTTGG CTTTCCAATT GCCAACCGCG AGAAAACTAA CTGTTGCAAG CAAAAATTTT TTTTGAATAC ACCGTGTATA ACAGTGCTAC
AAAGACTTGT AGGACTAATG AAACTCCAAG TTTCGGAACC GAAAGGTTAA CGGTTGGCGC TCTTTTGATT GACAACGTTC GTTTTTAAAA AAAACTTATG TGGCACATAT TGTCACGATG

      2050        2060        2070        2080        2090        2100        2110        2120        2130        2140        2150        2160
AACTAATACG CTCCCAACAT AATATACTAC TGTTAAATAC AACTGTTAAC AGATTTATCA GAATTGCTAT TTGGTAAAAC TGGCATTTTG GCAGCTGTGC ATAAATGGCC CATTAGCTCA
TTGATTATGC GAGGGTTGTA TTATATGATG ACAATTTATG TTGACAATTG TCTAAATAGT CTTAACGATA AACCATTTTG ACCGTAAAAC CGTCGACACG TATTTACCGG GTAATCGAGT
```

```
      2170        2180        2190        2200        2210        2220        2230        2240        2250        2260        2270        2280
GTTGGTTAGA GCGTCGTGCT AATAACGCGA AGGTCGCGGG TTCGATCCCC TCATGGGCCA GAATAATTTT TTTTTATTAC ATTTCTAATG CACTTTGAAA ATTAAAAAGT TAGTATAGAT
CAACCAATCT CGCAGCACGA TTATTGCGCT TCCAGCGCCC AAGCTAGGGG AGTACCCGGT CTTATTAAAA AAAAATAATG TAAAGATTAC GTGAAACTTT TAATTTTTCA ATCATATCTA

      2290        2300        2310        2320        2330        2340        2350        2360        2370        2380        2390        2400
TTGATGAGAA CGTGTCCCGC AAGCTTTTTG TACAAACCTA GATAAGCGAA GCCAATTATG GACGGCTGCT AATGTTATCT CATGCCGTGC ATGCTAGGAC AGTGAGTACC TGCGATGATG
AACTACTCTT GCACAGGGCG TTCGAAAAAC ATGTTTGGAT CTATTCGCTT CGGTTAATAC CTGCCGACGA TTACAATAGA GTACGGCACG TACGATCCTG TCACTCATGG ACGCTACTAC

      2410        2420        2430        2440        2450        2460        2470        2480        2490        2500        2510        2520
TCACATAAAG CCGACAAGGA TGATCTAACG AGGTCTATTT CTCGCAAGGC GGCTAAGTGA AATGTCAATT CAATTTACCA TCAAGGATGA TTGAATTTCC CGTCCCTTTA GCCGACCCAA
AGTGTATTTC GGCTGTTCCT ACTAGATTGC TCCAGATAAA GAGCGTTCCG CCGATTCACT TTACAGTTAA GTTAAATGGT AGTTCCTACT AACTTAAAGG GCAGGGAAAT CGGCTGGGTT

      2530        2540        2550        2560        2570        2580        2590        2600        2610        2620        2630        2640
CTTCTCCGCG TGTGAGGTAT CTAATTGCTT GGCAATCAAG CTAAACGGTT TATTGAACGA TCTTATCTTC GTAACGAGTT GAAAATGCTT TTAGACAATA TAATGGCAAA CAGAGAACAT
GAAGAGGCGC ACACTCCATA GATTAACGAA CCGTTAGTTC GATTTGCCAA ATAACTTGCT AGAATAGAAG CATTGCTCAA CTTTTACGAA AATCTGTTAT ATTACCGTTT GTCTCTTGTA

      2650        2660        2670        2680        2690        2700        2710        2720        2730        2740        2750        2760
CCTAATTTAT TGGTTCAACG CCCCATTTTG TATATACTTT TATGACACTA GATTTTGGGC GCGGAATCAG TGTCAGGATG GCCGAGCGGT CTAAGGCGCC AGACTCAAGA TTGAAAATCT
GGATTAAATA ACCAAGTTGC GGGGTAAAAC ATATATGAAA ATACTGTGAT CTAAAACCCG CGCCTTAGTC ACAGTCCTAC CGGCTCGCCA GATTCCGCGG TCTGAGTTCT AACTTTTAGA

      2770        2780        2790        2800        2810        2820        2830        2840        2850        2860        2870        2880
TACTTTCTGA ACGAAAGTGT TTGTTTAATG AGCGTTCTGG TCCTCTCTGA GGGCGTGGGT TCGAATCCCA CTTCTGACAA ATATTTTTTT AAGCCCAAAG AAATTAACAA TGATTTGACA
ATGAAAGACT TGCTTTCACA AACAAATTAC TCGCAAGACC AGGAGAGACT CCCGCACCCA AGCTTAGGGT GAAGACTGTT TATAAAAAAA TTCGGGTTTC TTTAATTGTT ACTAAACTGT

      2890        2900        2910        2920        2930        2940        2950        2960        2970        2980        2990        3000
AACAATGCAA AATTATTAAA TAATCAAATA TATATTTGCG TATACTTTTG AAAGATCTCT CTTTTGGGTT CACCTATATA AATACGTGAC CAATAAATAA AATGGTAACA AATACATTTG
TTGTTACGTT TTAATAATTT ATTAGTTTAT ATATAAACGC ATATGAAAAC TTTCTAGAGA GAAAACCCAA GTGGATATAT TTATGCACTG GTTATTTATT TTACCATTGT TTATGTAAAC

      3010        3020        3030        3040        3050        3060        3070        3080        3090        3100        3110        3120
CCATTGTATT AAATTAATAT ATAAATTCGT GACATATTTT TAGCAGGAAA ACGAAAAACG AAAAACGAAA ACTATCGTAG TAGCATTGAA ACATATAGTG TCAAATTTAG TTCTTATCAA
GGTAACATAA TTTAATTATA TATTTAAGCA CTGTATAAAA ATCGTCCTTT TGCTTTTTGC TTTTTGCTTT TGATAGCATC ATCGTAACTT TGTATATCAC AGTTTAAATC AAGAATAGTT

      3130        3140        3150        3160        3170        3180        3190        3200        3210        3220        3230        3240
TCCCTGAATC CAGATTTCAG TCAGGTTAAA AAAGTTGTAT ATTTTTATAT CAATATAAGC CATGTCGGAA GAGCAGCCAT CGAGCTCGTT GCTCAAGCCC ATTCTAATCA TGAAGGATAT
AGGGACTTAG GTCTAAAGTC AGTCCAATTT TTTCAACATA TAAAAATATA GTTATATTCG GTACAGCCTT CTCGTCGGTA GCTCGAGCAA CGAGTTCGGG TAAGATTAGT ACTTCCTATA
```

```
      3250        3260        3270        3280        3290        3300        3310        3320        3330        3340        3350        3360
CGATTTGGAC GGAGCAGGTG ATGGCATCGG TAAGGAGTCC TCGAAGACAA AGAAGCATGT AACCGTTCAA CCGGATGGTG GAGTATTATA CCTACTCGGA AACTTTGAGC AGCTGGAATA
GCTAAACCTG CCTCGTCCAC TACCGTAGCC ATTCCTCAGG AGCTTCTGTT TCTTCGTACA TTGGCAAGTT GGCCTACCAC CTCATAATAT GGATGAGCCT TTGAAACTCG TCGACCTTAT


      3370        3380        3390        3400        3410        3420        3430        3440        3450        3460
GGCGCCTTAT GTCCGACAAC GACTTGGTAG GTACAGAAAC ACCACAGGCA TGAATTCCCC ATTGATACGG TCTGTGGGCT TTTCAGCGCC AGAGCATGCC GA
CCGCGGAATA CAGGCTGTTG CTGAACCATC CATGTCTTTG TGGTGTCCGT ACTTAAGGGG TAACTATGCC AGACACCCGA AAAGTCGCGG TCTCGTACGG CT
```

Appendix II

Sequence Organization of Drosophila tRNA Genes

[Reprint from ICN-UCLA Symposium on Eucaryotic Gene Regulation, R. Axel and T. Maniatis, eds. (Academic Press, New York), 1979.]

EUCARYOTIC GENE REGULATION

SEQUENCE ORGANIZATION OF DROSOPHILA tRNA GENES[1]

Pauline Yen, N. Davis Hershey,
Randy R. Robinson, and Norman Davidson

Department of Chemistry, California Institute of Technology Pasadena, CA 91125

ABSTRACT The tRNA gene cluster at region 42A of chromosome arm 2R of the Drosophila melanogaster genome has been mapped on a number of recombinant DNA molecules with overlapping Drosophila inserts. The cluster extends over a region of 46 kb and contains at least 18 tRNA genes. It contains several $tRNA_2^{lys}$, $tRNA_2^{arg}$, and $tRNA^{asn}$ genes that are mutually interspersed and irregularly spaced. A preliminary account of experience with Benton-Davis screening for λ plaques with inserts carrying tRNA genes is given.

INTRODUCTION

Saturation hybridization experiments indicate that there are 600-800 tRNA genes in the haploid genome of Drosophila melanogaster. (1,2,3) There are about 90 resolvable peaks when total Dm tRNA is fractionated by RPC-5 chromatography(4). From an analysis of the kinetics of tRNA-genomic DNA hybridization, Weber and Berger estimate that Dm tRNA is made up of about 59 kinetic families (3). These data suggest that there are approximately 60-90 different tRNA sequences (*i.e.*, species) in Drosophila and an average reiteration frequency for any one species of six to 13.

Steffensen and Wimber (5) and later Elder, Szabo and Uhlenbeck (6) have mapped the sites on polytene chromosomes to which total labeled Dm tRNA hybridizes. In the latter study, 54 sites were mapped by hybridization of ^{125}I-labeled tRNA to polytene chromosomes. Of these 26 were estimated by grain counts to be strong sites with 4 or more gene copies, and 28 were weak with 1-3 gene copies. These sites are listed in Table 1.

The authors noted that other sites, such as those with small numbers of tRNA genes, those with tRNA genes of low

[1]This work was supported by a grant GM 10991 from the United States Public Health Service.

ISBN 0-12-068350-4

abundance, or those on portions of the genome that are under-replicated in polytene chromosomes (*e.g.*, the Y chromosome) may not have been detected.

Several laboratories have carried out *in situ* studies on purified tRNAs. Some of these results are also reported in Table 1 (7,8,9,10).

The overall conclusions from these studies are that most of the tRNA genes of Drosophila are clustered at a limited

Table 1

Drosophila 4S RNA Sites[a]

X	2L	2R	3L	3R	4
3D	22DE	41CD	61D Met3	84A Met2	none
3F	23EF	*42A Asn Arg2, Lys2	*62A Lys2 Glu4	84D Val3b	
5F6A	24DE	42E Lys2	63A	84F Arg2	
11A	*27D	*43A	63B	87BC	
*12E	28D	44EF	64DE Val3a	87F88A	
	28F29A	46A Met3	66B	89BC Val4	
	29D Asp2	48AB Met2	67B	90C Val3b	
	29E	48CD	69F70A	90DE	
	*34A	49AB	70BC Val4	*91C	
	35AB	49F50A	70DE	92AB Val3b	
		*50BC Lys2	72F*73A Met2	93A	
		52F53A Glu4	79F	95F96A	
		54A		97CD	
		54D		99EF	
		55EF			
		56D Val4			
		*56F Glu4			
		57D			
		58AB			
		60E			

[a]The locus assignments in heavy type were made by Elder et al. (6) using total labeled tRNA. Sites that were heavily labeled are underlined. Entries to the side of the main columns of the table are assignments made using purified tRNA's (5,6,7,8,9,10). Asterisks indicate assignments for recombinant Charon plaques that give strong tRNA signals, isolated in our laboratory. Entries in italics are *in situ* assignments that may be different from those of Elder *et al.* (6).

number of sites, perhaps 54 to 60, on the polytene chromosomes; that some of these clusters contain more than one species of tRNA gene; and that for some tRNA species there are genes at 2 or more sites.

We have previously described our initial studies of the sequence arrangement of the tRNA genes on a cloned fragment of Dm tRNA of length 9 kb (11). This fragment was shown by *in situ* hybridization to map at region 42A of chromosome 2R. *In situ* studies with labeled total Dm tRNA had shown that 42A is one of the richest tRNA gene clusters in the Drosophila melanogaster genome (5).

In the present communication, we describe the present status of our further studies of the sequence organization of tRNA genes in cloned DNA derived from this region of the chromosome. We also give a preliminary account of our experience in screening for recombinant λ plaques with Drosophila tRNA genes.

EXPERIMENTAL

A recombinant "library" of nuclear Dm DNA inserted into the modified λ bacteriophage Charon 4 (12) has been constructed. High molecular weight nuclear Dm DNA was partially digested with EcoR1 to varying extents (5%, 7.5%, 14%, 20%, 31% and 57%) of total digestion. Fractions of length 22 ± 4 kb were isolated from each digest by sucrose gradient velocity sedimentation. Equal masses of sized DNA from each digest were combined and then ligated to annealed EcoR1 ends of Charon 4 DNA. The ligation products were packaged *in vitro* (13,14) to yield 2×10^6 recombinant plaque forming units.

The library contains inserts ranging in size from 12 to 20 kb, and contains less than 1% Charon 4 contamination. Library Dm DNA is representative of the single copy Dm genome as demonstrated by its ability to drive 95% of single copy tracer into duplex with a $cot_{\frac{1}{2}}$ of 280, compared to 270 for nuclear Dm DNA driver.

The other experimental methods used here are all standard, and are referred to where necessary in the Results section.

RESULTS

a) The Gene Cluster at 42 A. The plasmid pCIT12 described in our initial study (11) contains a Dm DNA segment of length 9.3 kb fused to the vector ColE1. Fig. 1 includes a map of the Eco R1 cleavage sites and identified positions of tRNA genes on this insert. The methods by which the tRNA genes have been mapped and identified will be discussed later.

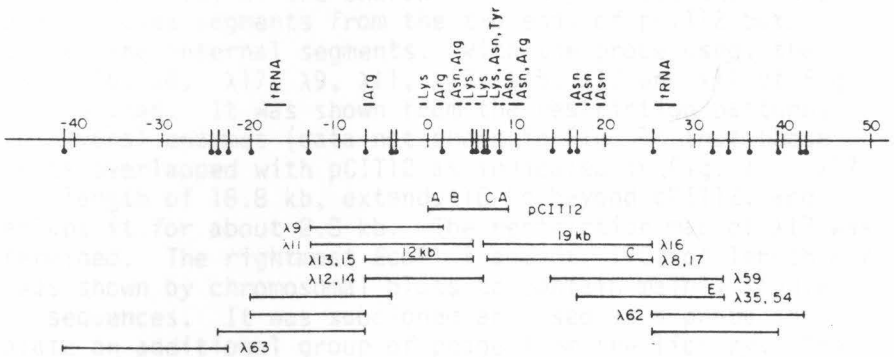


FIGURE 1. Dm inserts in Charon 4 phage covering the region 42A. The scale is in kilobases (kb). ↓ denote EcoR1 clevage sites. Capital letters denote EcoR1 restriction fragments referred to in the text. Note that fragment A of pCIT12 contains the Col E1 vector as well as segments on both ends of the insert. Positions of tRNA genes were determined by blotting experiments with labeled total or purified tRNA probes as described in the text.

It seemed probable from the strong *in situ* hybridization of total Dm tRNA to locus 42A that the 9.3 kb segment of pCIT12 does not include the entire gene cluster at locus 42A. This inference was confirmed by our initial studies of Charon 4 recombinant molecules with inserts that overlapped with pCIT12, and by the studies of P. Gergan and P. Wensink who have independently isolated a Dm plasmid which partially overlaps with pCIT12 but extends further to the right (given the arbitrary orientation of the map of Fig. 1) and includes additional tRNA genes. We therefore wished to determine the total extent of the tRNA gene cluster of band 42A which includes pCIT12.

For this purpose we have engaged in the procedure of "walking down the chromosome". Initial attempts to isolate recombinant molecules with adjacent sequences from the Charon 4 library using pCIT12 DNA as a probe gave many plaques with inserts derived from other regions of the genome. This was due to the presence of repeated sequences in the EcoR1 fragment C of pCIT12. Therefore, a DNA segment consisting of the 2 R1 segments A and B of pCIT12 (which in addition to the Dm segment of the A fragment indicated in Fig. 1 includes the

6.7 kb of the vector ColE1) was subcloned. This DNA was labeled by nick translation and used as a probe in a Benton-Davis screen (15) of the Charon 4 library. Note that this probe includes segments from the two ends of pCIT12 but excludes the internal segments. With the probe used, the phages λ16, λ8, λ17, λ9, λ11, λ13, λ15, λ12 and λ14 of Fig. 1 were isolated. It was shown from the restriction patterns with several enzymes (data not shown in Fig. 1) that these inserts overlapped with pCIT12 as indicated in Fig. 1. λ17 has a length of 18.8 kb, extends 16 kb beyond pCIT12, and overlaps it for about 2.8 kb. The restriction map of λ17 was determined. The rightmost EcoR1 fragment, 17C, of length 8.7 kb was shown by chromosomal blots to contain mainly single copy sequences. It was subcloned and used as a probe to isolate an additional group of phage from the library. The restriction patterns of these phage were determined. Those that did not contain the EcoR1 fragments 17D and 17F were selected for further study. These include λ59, λ35, and λ54. These three phage all end at a point 8 kb to the right of λ17. The process was repeated using the 1.6 kb fragment 35E and phage extending to the right selected. The phage λ62 extends 9 kb beyond λ35.

In the same way phage extending to the left of pCIT12 were selected. The phage λ11 has a Dm DNA insert of length 18.4 kb and extends 13.3 kb beyond pCIT12. In a series of two additional steps which are obvious by inspection of Fig. 1, the phage λ63 was isolated. Its end point is about 41 kb beyond the left end of pCIT12. Thus, the set of recombinant molecules cover a region of length 83 kb at locus 42A.

Preliminary tRNA gene assignments on some of the cloned segments have been made by Southern blotting experiments, using several different kinds of probes. Total Dm 4S RNA has been purified from Drosophila pupae as previously described (11) and labelled with ^{32}P either at the 5' end by T4 polynucleotide kinase or at the 3' end as (5'-^{32}P)pCp with T4 RNA ligase (16). The specific activities achieved either way are approximately 2×10^7 cpm/μg tRNA. However, the tRNA is less degraded by the 3' end labeling procedure.

Specific genes have been identified with probes prepared by labeling purified Dm tRNA species. These latter preparations were generous gifts from the laboratory of Dr. Dieter Söll. Since the tRNAs are not always completely purified and contain other tRNA components as minor contaminants, and since the hybridizations are carried out in vast tRNA excess, care must be taken to carry out hybridizations for times such that only the major species gives strong spots in the autoradiographs. Weak signals which are probably due to minor tRNA impurities have been disregarded in the tRNA gene assignments

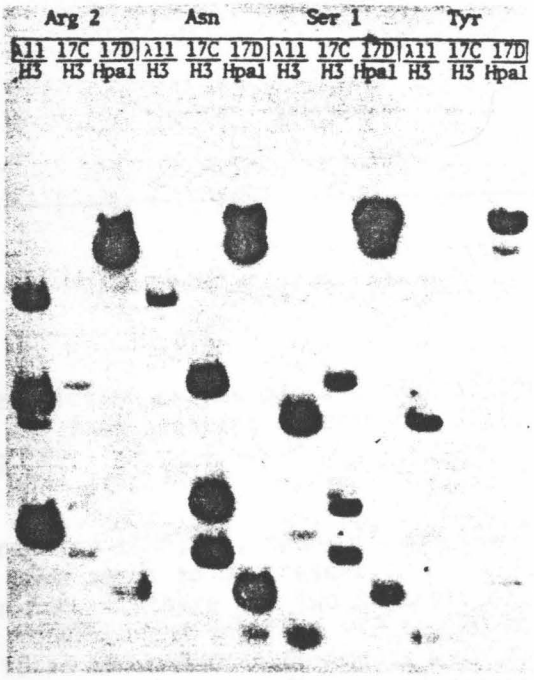


FIGURE 2. Blotting experiments with purified tRNA probes on appropriate restriction digests of the Eco R1 fragments C and D of λ17 (see Fig. 3) recloned in pBR322, as well as of a Hind III digest of λ11. The $tRNA^{ser}$ probe was heavily contaminated with $tRNA_2^{lys}$ and further experiments with purified $tRNA_2^{lys}$ show that the positive signals are due to this tRNA (D. Söll, personal communication). Because of impurities in the tRNAs, only strong autoradiographic signals are interpreted as positive for the main tRNA component.

in Figs. 1 and 3.

An example of such a Southern blot experiment with several purified tRNA species is given in Fig. 2. The tRNA gene assignments given in Figs. 1 and 3 have been made, on the basis of experiments like those of Fig. 2 in our laboratory, in that of Dr. Dieter Söll (17), and by P. Gergan and P. Wensink (personal communication).

Figs. 1 and 3 show that the gene cluster at region 42A on chromosome arm 2R of the Dm genome contains multiple

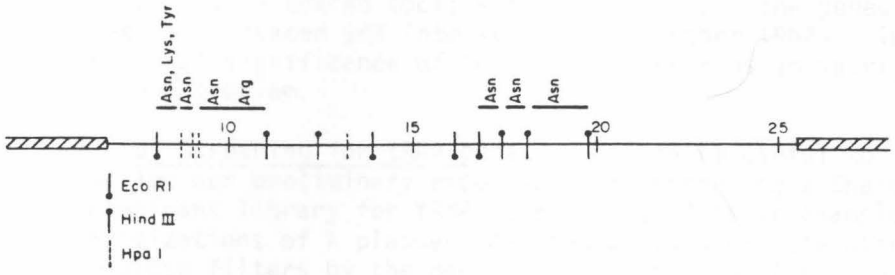


FIGURE 3. Restriction enzyme map and tRNA assignments for the DNA phage λ17 from blotting experiments such as illustrated in Fig. 2.

copies of tRNA genes for $tRNA_2^{arg}$, $tRNA_2^{lys}$, and $tRNA^{asn}$. At least one $tRNA^{tyr}$ gene seems to be present.

On the map in Fig. 1, there are two segments of DNA, one extending from 25 to 28 kb and one from -17 to -20 kb, which hybridize with a total RNA probe; experiments with specific tRNAs have not yet been performed. The hybridization experiments are not quantitative and the possibilities exist that: a) several genes for a particular species are present on a restriction fragment where only one has been assigned; b) there is a restriction endonuclease site sufficiently close to the center of a gene so that two restriction fragments hybridize to the tRNA probe even though there is only one gene; c) other as yet unidentified tRNA species, for which purified probes were not available, are also present.

The map of pCIT12 (11,17, P. Gergan and P. Wensink, personal communication) and the present mapping studies of λ17 (Fig. 3) show that the several species of genes are interspersed and irregularly spaced. At the level of resolution achieved at present, there is no evidence of extensive sequence homology in the regions flanking identical genes.

It is apparent from these studies that there is a segment of DNA of length 47 kb in the region 42A on chromosome 2R of the Dm genome which contains at least 18 tRNA genes. For a region of at least 14 kb on the right and 23 kb on the left there are no additional tRNA genes so that it is possible that this segment represents the complete tRNA gene cluster at region 42A.

By *in situ* hybridization experiments, Tener *et al.* (7) have estimated that there are approximately 18 $tRNA_2^{lys}$ genes

in the haploid genome of Dm, and that these map at the sites 42A, 42E, 50B, 62A and 63B. The $tRNA_2^{arg}$ genes have been assigned to chromosomal locations of 42A and 84F.

It thus appears that at least for some of the repeated tRNA gene families of Drosophila, genes are located at several widely spaced loci; within each locus, the genes are irregularly spaced and interspersed with other tRNAs. The functional significance of this pattern remains an intriguing, unsolved problem.

b) Screening for tRNA genes. It may be useful to describe our preliminary experiences in screening a Charon 4 recombinant library for tRNA genes. Fig. 4 is an example of hybridizations of λ plaques adsorbed on to duplicate nitrocellulose filters by the Benton-Davis procedure (15). Hybridizations were carried out with 5' ^{32}P labeled total pupal tRNA. The plate in Fig. 4 contains about 1700 plaques. Of these, 29 gave apparent duplicate positive signals. These positive plaques were picked and replated. Seventeen gave

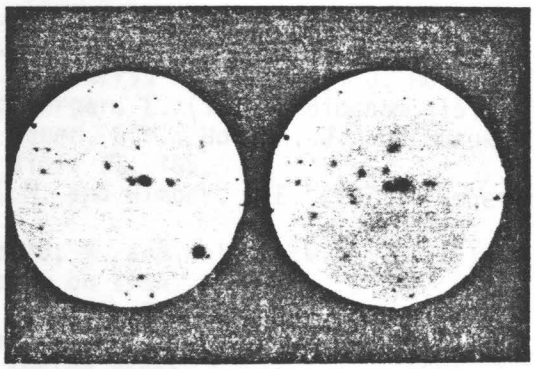


FIGURE 4. Duplicate Benton-Davis (15) screens of the Charon 4 Drosophila library. Each plate contained about 1700 plaques. Hybridizations with 5' ^{32}P labeled tRNA were conducted with 0.15 μg/ml tRNA ($\sim 2 \times 10^7$ cpm/μg) for 15 hrs at 68°C in 1.0 M NaCl, 0.050 M Tris, pH 8; 1 mM EDTA, 10 x Denhardt's solution, 0.1% SDS, 0.1% sodium pyrophosphate, and 20 μg/ml Dm rRNA.

positive plaques on at least one rescreen; thirteen were positive on duplicate rescreens. Nine of the strongest spots were selected for further study.

Similar results were obtained in screening the entire library (10 plates). Comparable results were obtained using ^{125}I labeled tRNA probes. A number of the plaques which gave strong signals were purified and characterized by hybridization, with the results recorded in Table 1.

At the present time, we do not know the sensitivity of the screening procedure. We suspect however that, with improvements in labeling and hybridization procedures, plaques with 3 or 4 tRNA genes can be identified.

ACKNOWLEDGMENTS

We are indebted to Drs. Dieter Söll, Olke Uhlenbeck, and Pieter Wensink for several personal communications. The purified tRNA's used were a generous gift from the laboratory of Dr. Dieter Söll. Department of Chemistry Contribution #5997.

REFERENCES

1. Ritossa, F.M., Atwood, K.C. and Spiegelman, S. (1966). Genetics. 54, 663.
2. Tartof, K.D. and Perry, R.P. (1970). J. Mol. Biol. 98, 503.
3. Weber, L. and Berger, E. (1976). Biochem. 15, 5511.
4. White, B.N., Tener, G.M., Holden, J. and Suzuki, D.T. (1973). Dev. Biol. 33, 185.
5. Steffensen, D.M. and Wimber, D.E. (1971). Genetics. 69, 163.
6. Elder, R., Szabo, P. and Uhlenbeck, O. (1979). Cold Spring Harbor Symposium on tRNA. in press.
7. Tener, G.M., Hayashi, S., Dunn, R., Delaney, A., Gillam, I.C., Grigliatti, T.A., Kaufman, T.C. and Suzuki, D.T. (1979). Cold Spring Harbor Symposium on tRNA. in press.
8. Dunn, R., Hayashi, S., Gillam, I.C., Delaney, A.D., Tener, G.M., Grigliatti, T.A., Kaufman, T.C. and Suzuki, D.T. (1979). J. Mol. Biol. in press.
9. Kubli, E. and Schmidt, T. (1978). Nucl. Acid Res. 5, 1465.
10. Schmidt, T., Egg, A.H. and Kubli, E. (1979). Molec. Gen. Genet. in press.
11. Yen, P.H., Sodja, A., Cohen, M., Conrad, S.E., Wu, M., Davidson, N. and Ilgen, C. (1977). Cell. 11, 763.
12. Blattner, F.R., Williams, B.G., Blechi, A.E., Thompson, K.D., Faber, H.E., Furlong, L.A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L. and Smithies, O. (1977). Science. 196, 161.

13. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978). Cell. 15, 687.
14. Sternberg, N., Tiemeier, D. and Enquist, L. (1977). Gene. 1, 255.
15. Benton, W.D. and Davis, R.W. (1977). Science. 196, 180.
16. England, T.E. and Uhlenbeck, O.C. (1978). Nature. 275, 560.
17. Schmidt, O., Mao, J.-I., Silverman, S., Hovemann, B. and Söll, D. (1978). Proc. Natl. Acad. Sci. USA. 75, 4819.

APPENDIX III

Recombinant DNA Studies of DNA Sequence Organization Around Actin and tRNA Genes of Drosophila melanogaster

[Reprint from RNA Polymerase, tRNA and Ribosomes: Their Genetics and Evolution, S. Osawa, H. Ozeki, H. Uchida, and T. Yura, eds. (University of Tokyo Press), in press.]

RECOMBINANT DNA STUDIES OF DNA SEQUENCE ORGANIZATION AROUND ACTIN AND tRNA GENES OF DROSOPHILA MELANOGASTER

Norman Davidson
Eric A. Fyrberg
N. Davis Hershey
Karen Kindle
Randy R. Robinson
Ann Sodja*
Pauline Yen

Department of Chemistry
California Institute of Technology
Pasadena, California 91125 USA

*Department of Biology
Wayne State University
Detroit, Michigan 48202 USA

Running Title: Actin and tRNA Genes of Drosophila

Abstract

The construction and characterization of a Drosophila recombinant DNA library with inserts of length 12-20 kb in the lambda vector Charon 4 is described. The library represents 95 ± 5% of the Drosophila genome. A recombinant clone containing a Drosophila actin gene has been selected and mapped. The actin gene consists of a 5' proximal exon of length 60-100 nucleotides, an intervening sequence (intron) of length 1.6 kb, and a main exon of length 1.6 kb. The tRNA gene cluster around locus 42A of the Drosophila chromosome has been characterized. There are at least 14 tRNA genes in a region of length 50 kb, flanked on each site by segments of length 20 kb with no tRNA genes. At least 5 different tRNA species are present, and they are distributed in an irregular, interspersed pattern.

Introduction

We wish to describe several applications of recombinant DNA methodology to study the organization of expressed genes and their flanking sequences in the DNA of a higher eukaryote. The examples considered are the actin genes and one particular subset of tRNA genes in Drosophila melanogaster. We also discuss the problem of preparing a set of recombinant clones which includes almost all of the sequences in the Drosophila genome.

The Drosophila Library

If random fragments of the genomic DNA from a particular organism are cloned in a bacterial host, and if the number of fragments is sufficient so that a substantial fraction of the total genome of the organism is included, this collection of clones is called a "library" (1).

A library of Drosophila recombinant DNA molecules cloned in the vector Charon 4 (2) has been prepared. This particular vector has a cloning capacity of 8 to 22 kb. The procedure for preparing the library is depicted in Fig. 1.

Fig. 1

In brief, high molecular weight nuclear Dm DNA was partially digested with EcoR1 to varying extents (5%, 7.5%, 14%, 20%, 31%, and 57%) of total digestion. Fractions of lengths 22 $\pm$ 4 kb were isolated from each digest by sucrose gradient velocity sedimentation. Equal masses of size fractionated DNA from each digest were combined and then ligated to cohered EcoR1 arms of Charon 4 DNA. The ligation products were packed *in vitro* (3,4) to yield 2×10^6 recombinant plaque forming units.

In spite of the effort to obtain only larger inserts by the size fractionation, we observe that the library actually contains inserts ranging in length from 12-20 kb. Less than 1% of plaque forming phage do not contain Drosophila inserts.

Given a suitable radioactive probe for a given gene, bacteriophage libraries can be successfully screened for those recombinant DNA molecules containing the gene by the rapid *in situ* plaque hybridization procedure of Benton and Davis (5). These technical accomplishments are described in detail by Maniatis et al., (1).

A question of crucial important if one wishes to find a particular gene is that of the representation of the library, that is, the fraction of all sequences from the organisms's genome that is present in the library DNA. This problem admits a precise solution if the insert fragments are prepared by perfectly random shear. For this case, let L be the average length of the insert, G the haploid genome length, and N the number of independently formed bacteriophage in the library. The probability of not finding a given nucleotide in the library is then simply given as

$$[1-(L/G)]^N \approx \exp(-NL/G)$$

Thus, if L ≈ 18 kbp (kilobase pairs) as for Charon 4 and G = 1.8×10^5 kbp as for Drosophila, approximately 40,000 independently derived bacteriophage should be screened to give a 98% probability of finding a particular sequence in the library.

The extent to which an endonuclease partial digest library mimics a random shear library is not known. This probability depends on the extent and randomness of partial digestion, on the minimum insert length which is to be cloned, and on the length of the cloning window, that is the difference in length between the maximum and minimum inserts which can be cloned. This problem has been studied theoretically but a reliable numerical approximation to the theoretical equations is not available (6).

A brief discussion of our experimental test of the representation of the library and of the parameters which, according to theory, influence the representation is given later in this article.

The Actin Genes

Actin is a major protein in both muscle and non-muscle cells. In vertebrates, protein sequence studies show that there are at least 6 distinct polypeptide sequences for actin extracted from skeletal muscle, cardiac muscle, smooth muscle (2 sequences) and non-muscle (2 sequences) cells (7). There must be at least one gene, but possibly more, coding for each actin polypeptide. In Dictyostelium the reiteration frequency of actin gene sequences has been estimated at 15-20, and there appear to be at least 2 different polypeptide sequences (8).

Actin is an abundant protein in both muscle and non-muscle cells. However, unlike the globin and ovalbumin genes of vertebrates, there is no tissue or system in Drosophila from which it is practical to purify actin mRNA in sufficient purity to use it directly as a probe for selecting actin genes from the recombinant library. However, the similarities and properties in amino acid composition of actin throughout the animal kingdom suggest that it is a highly conserved protein (9). Therefore it seemed probable that an actin nucleic acid probe isolated from some other animal would be useful for selecting Drosophila actin genes.

The procedure that was used for isolating a Drosophila actin gene from the Charon 4 library was as follows. Total poly A^+ RNA was extracted from Drosophila larval cuticles in connection with a screen for abundantly expressed genes, in general. The larval cuticle cells contain significant amounts of

actin protein and therefore presumably of actin mRNA. It was anticipated that actin mRNA would be present at a level of 0.5 -5% in such a preparation. A ^{32}P-labeled cDNA probe was prepared from this RNA, and used to perform a Benton-Davis *in situ* plaque hybridization screen (5) on 25,000 plaques from the library. Many plaques which gave positive signals of detectable intensity were observed; 20 which gave very strong signals were selected. The above screening procedure was carried out by Dr. Jay Hirsh. Phage were plaque purified from these 20 plaques and rescreened using a poly A^+ mRNA preparation from chick embryo muscle that had been enriched for actin mRNA by sucrose velocity centrifugation. Three positive plaques were then selected. One of these gave a positive signal with a probe prepared from the cloned Dictyostelium actin gene (8). This bacteriophage contains the Drosophila actin λDmA2 recombinant molecule described below.

A map of the Drosophila insert on the cloned DNA of λDmA2 is given in Fig. 2. The positions of various restriction endonuclease sites, as well as of the sequences complementary to actin mRNA, are depicted.

Fig. 2

The actin gene has a total length of 3.3 kb. It consists of two regions coding for sequences present in actin mRNA (that is, two exons) separated by an intervening sequence (an intron) which is not present in the mature mRNA. The 5' proximal exon is a very short leader sequence of estimated length 60-100 nt. This is followed (to the right in Fig. 2) by a 1.6 kb intron, which is in turn followed by the 1.6 kb exon which constitutes the main body of the actin mRNA. Actin polypeptides have a total length of approximately 375 amino acids. Thus the translated sequences in the mRNA are about 1.1 kb in length. The remaining sequences in the 1.7 kb mRNA are untranslated.

The map in Fig. 2 has been established by electron microscopic mapping of R loop hybrids formed by hybridization of actin mRNA (present in total Drosophila

RNA) to λDmA2 DNA under appropriate conditions (10). Several such micrographs demonstrating the presence of an intervening sequence and the RNA:DNA hybrid region, are shown in Fig. 3.

Fig. 3

In addition to the electron microscopic mapping, the position of the main coding sequence within the two central Sal I fragments shown in Fig. 2 has been demonstrated by the now standard technique of Southern (11) - that is, gel electrophoresis of a restriction digest followed by transfer of the denatured DNA to nitrocellulose filters, and hybridization with suitable probes. In this case, the probes used have been either cDNA from total poly A^+ larval RNA or the nick translated Dictyostelium actin plasmid.

Fig. 4

Figure 4 illustrates an application to this particular gene of a method for assigning the 5' to 3' polarity of an RNA by restriction endonuclease digestion, gel electrophoresis, Southern blotting, and hybridization with suitable probes. The method depends on the fact that cDNA made from mRNA by reverse transcription with an oligo (dT) primer has a preponderance of sequences from the 3' end of the RNA whereas cDNA made with a random calf thymus primer (12) has a uniform representation of all of the sequences in the RNA.

The principle is that if the coding part of a gene extends across a restriction endonuclease site (in this case a Sal I site), then the ratio of hybridization intensities of the segment containing the 3' end of the gene to that of the segment containing the 5' end of the gene will be lower for the random primed cDNA than for the oligo (dT) primed cDNA. Further details of the experiment are explained in the legend to the figure.

The number of actin genes in the genome has been determined by hybridizing Southern blots of genomic DNA digested by various restriction enzymes to actin probes (for example, the 1.8 kb Hind III fragment shown in Fig. 2 that contains the bulk of the actin coding region of λDmA 2). When such digests are compared

to control lanes in which genome equivalents of the cloned restriction fragments have been electrophoresed, it becomes clear that there are 6 actin genes in the haploid Drosophila genome.

These 6 distinct genes are located at 5 sites on the polytene chromosome as determined by in situ hybridization (13). These sites are: 5C on the X chromosome, 57A on 2R, 87E and 88E on 3R, and 79B on 3L. Present evidence indicates that, at the position 5C on the X chromosome, there are actually 2 distinct genes which are close enough to each other so that they are not resolved at the level of in situ hybridization (that is 100-200 kb). However, preliminary analysis of cloned DNAs indicates that these two genes are separated by greater than 10 kb.

Thus, we have shown at the present state of our studies that there are 6 distinct actin genes at 6 distinct chromosomal sites. We have not yet established whether or not these 6 genes differ in the amino acid sequences for which they code. However, there are several different actin mRNAs of different molecular lengths in Drosophila (our preliminary results) just as in vertebrates (14). There are several electrophoretic variants of the protein in Drosophila (15,16). It is therefore a plausible hypothesis that in Drosophila, as in vertebrates, there are different nucleic acid sequences coding for several different muscle types of actin and for one or several non-muscle actins.

Thus, the actin genes are an interesting multigene system for further investigation. In a particular type of muscle tissue, it may be presumed that other specific muscle protein genes will be coordinately expressed with a particular actin gene. Identification of the structural factors which regulate this coordinate expression is a problem remaining for future studies. On the other hand, different actin genes expressed at different stages of development must be controlled by different regulatory elements.

Drosophila tRNA Genes

Saturation hybridization experiments indicate that there are 600-800 tRNA in the haploid genome of Drosophila melanogaster (17,18,19). There are about 90 resolvable peaks when total Dm tRNA is fractionated by RPC-5 chromatography (20). From an analysis of the kinetics of tRNA-genomic DNA hybridization, Weber and Berger estimate that Dm tRNA is made up of about 59 kinetic families (19). These data suggest that there are approximately 60-90 different tRNA sequences (i.e., species) in Drosophila and an average reiteration frequency for any one species of six to 13.

Steffensen and Wimber (21) and later Elder, Szabo and Uhlenbeck (22) have mapped the sites on polytene chromosomes to which total labeled Dm tRNA hybridizes. In the latter study, 54 sites were mapped by hybridization of ^{125}I-labeled tRNA to polytene chromosomes. Of these 26 were estimated by grain counts to be strong sites with 4 or more gene copies, and 28 were weak with 1-3 gene copies. These sites are listed in Table 1. The authors noted that other sites, such as those with small numbers of tRNA genes, those with tRNA genes of low abundance or those on portions of the genome that are underreplicated in polytene chromosomes (e.g., the Y chromosome) may not have been detected.

Several laboratories have carried out *in situ* studies on purified tRNAs. Some of these are also reported in Table 1 (23,24,25,26,27).

Table 1

The overall conclusions from these studies are that most of the tRNA genes of Drosophila are clustered at a limited number of sites, perhaps 54 to 60, on the polytene chromosomes; that some of these clusters contain more than one species of tRNA gene; and that for some tRNA species there are genes at 2 or more sites.

We have previously described our initial studies of the sequence arrangement of the tRNA genes on a cloned recombinant plasmid pCIT 12 (28). This plasmid contains a Drosophila DNA insert of length 9.3 kb fused to the vector Col E1.

A map of the distribution of restriction sites and tRNA genes on this insert is shown in Fig. 5. In the original study, the tRNA genes were mapped by electron microscopy of ferritin labeled tRNA and by the Southern blotting, hybridization procedure using total ^{3}H labeled tRNA as a probe. These studies revealed that the insert contains 3 genes that are either identical or closely related in sequence, labeled tRNA-2, tRNA-3, and tRNA-4 and a different gene, tRNA-1, all as shown in Fig. 5.

Fig. 5

When single strands of the DNA were mounted for electron microscopy, several secondary structure features consisting of a duplex stem, usually separated by a single-stranded loop, were observed at reproducible positions. The positions of the stem sequences and their inverted complements are marked as c, c'; a, a', etc., in Fig. 5.

More refined mapping has been accomplished by hybridization of partially purified individual tRNA probes (29,30; P. Wensink, P. Gergan, Brandeis University, personal communication) and most decisively by direct DNA sequencing (30). These studies show that there are actually 9 tRNA genes on the Drosophila insert of pCIT12, as indicated in Fig. 5.

The sequencing data show that the several inverted repeat segments are actually tRNA genes present in two copies on opposite strands. Obviously, the presence of a pair of closely spaced complementary DNA segments that can hybridize with each other on a single strand of DNA complicates the problem of one segment hybridizing to a complementary tRNA probe.

The insert on pCIT12 was shown by in situ hybridization to map at region 42A of chromosome 2R. In situ studies with labeled total Drosophila tRNA had shown that 42A is one of the richest tRNA gene clusters in the Drosophila melanogaster genome (21). It therefore seemed probable that the 9.3 kb segment of pCIT12 does not include the entire gene cluster at locus 42A. This inference was confirmed by our initial studies of Charon 4 recombinant molecules with inserts that overlapped with pCIT12, and by the studies of P. Gergan and P. Wensink who have independently isolated a Drosophila plasmid which partially overlaps with pCIT12 but extends further to the right (given the arbitrary orientation of the map of Fig. 5) and includes additional tRNA genes. We therefore wished to determine the total extent of the tRNA gene cluster of band 42A which includes pCIT12.

For this purpose we have engaged in the procedure of "walking down the chromosome". Initial attempts to isolate recombinant molecules with adjacent sequences from the Charon 4 library using pCIT12 DNA as a probe gave many plaques with inserts derived from other regions of the genome. This was due to the presence of repeated sequences in the EcoR1 fragment C of pCIT12. Therefore, a DNA segment consisting of the EcoR1 segments A and B of pCIT12 (which in addition to the Dm segment of the A fragment indicated in Fig. 5 includes the 6.7 kb of the vector ColE1) was subcloned. This DNA was labeled by nick translation and used as a probe in a Benton-Davis screen (5) of the Charon 4 library. Note that this probe includes segments from the two ends of pCIT12 but excludes the internal segments. With the probe used, the phages λ16, λ8, λ17, λ9, λ11, λ13, λ15, λ12 and λ14 were isolated. It was shown from the restriction patterns with several enzymes (data not shown) that these inserts overlapped with pCIT12 as indicated in Fig. 6. λ17 has a length of 18.8 kb, extends 16 kb beyond pCIT12, and overlaps it for about 2.8 kb. The

Fig. 6

restriction map of λ17 was determined. The rightmost EcoR1 fragment, 17C, of length 8.7 kb was shown by chromosomal blots to contain mainly single copy sequences. It was subcloned and used as a probe to isolate an additional group of phage from the library. The restriction patterns of these phage were determined. Those that did not contain the EcoR1 fragments 17D and 17F were selected for further study. These include λ59, λ35, and λ54. These three phage all end at a point 8 kb to the right of λ17. The process was repeated using the 1.6 kb fragment 35E and phage extending to the right selected. The phage λ62 extends 9 kb beyond λ35. The right most R1 fragment of λ62 was used as probe and λ85, 86 and 88 were selected. The phage λ88 extends 10 kb beyond λ62.

In the same way phage extending to the left of pCIT12 were selected. The phage λ11 has a Dm DNA insert of length 18.4 kb and extends 13.3 kb beyond pCIT12. In a series of two additional steps which are obvious by inspection of Fig. 1, the phage λ63 was isolated. Its end point is about 41 kb beyond the left end of pCIT12. Thus, the set of recombinant molecules cover a region of length 94 kb at locus 42A. Preliminary tRNA gene assignments on some of the cloned segments have been made by Southern blotting experiments, using several different kinds of probes. Total Dm 4S RNA has been purified from Drosophila pupae as previously described (28) and labeled with ^{32}P either at the 5' end by T4 polynucleotide kinase or at the 3' end as (5'-^{32}P)pCP with T4 RNA ligase (31). The specific activities achieved either way are approximately 2 x 10^7 cpm/μg tRNA. However, the tRNA is less degraded by the 3' end labeling procedure.

Specific genes have been identified with probes prepared by labeling purified Dm tRNA species. These latter preparations were generous gifts from the laboratory of Dr. Dieter Söll. Since the tRNAs are not always

completely purified and contain other tRNA components as minor contaminants, and since the hybridizations are carried out in vast tRNA excess, care must be taken to carry out hybridizations for times such that only the major species gives strong spots in the autoradiographs. Weak signals which are probably due to minor tRNA impurities have been disregarded in the tRNA gene assignments.

The hybridization experiments are not quantitative and the possibilities exist that: a) several genes for a particular species are present on a restriction fragment where only one has been assigned; b) there is a restriction endonuclease site sufficiently close to the center of a gene so that two restriction fragments hybridize to the tRNA probe even though there is only one gene; c) other as yet unidentified tRNA species, for which purified probes were not available, are also present. On the map in Fig. 6, there is a segment of DNA extending from -17 to -20 which hybridizes with a total tRNA probe, however, it does not hybridize with purified tRNA lys, arg or asn probes. Therefore it contains different tRNA's

By in situ hybridization experiments, Tener et al. (23) have estimated that there are approximately 18 $tRNA_2^{lys}$ genes in the haploid genome of Drosophila and that they map at the sites 42A, 42E, 50B, 62A and 63B. The $tRNA_2^{arg}$ genes have been assigned to chromosomal locations of 42A and 84F.

It thus appears that at least for some of the repeated tRNA gene families of Drosophila, genes are located at several widely spaced loci; within each locus, the genes for any one tRNA species are irregularly spaced, occur on both strands, and are interspersed with other tRNA genes. The functional significance of this pattern remains an intriguing, unsolved problem.

One further interesting result, bearing on the representation problem disucssed in the next section, is shown in Fig. 6. The bacteriophage with

numbers below 100 that extend either to the left or to the right of pCIT12 were all discovered in the Charon 4 partial R1 library. However, no bacteriophage with an insert extending both to the left and the right of the 9 kb insert of pCIT12 was discovered in spite of extensive screening. We therefore screened the random shear library prepared by J. Lauer (1). This resulted in the discovery of the phages with 100 numbers spanning this region, as shown in Fig. 6.

The Representation Problem

We wished to determine as accurately as possible what fraction of the total Drosophila genome is present in our partial R1 library. The problem was studied by the methods described by Galau et al. (32). We first prepared single copy Drosophila nuclear DNA by suitable cycles of reassociation and HAP chromatography. This DNA was labeled with ^3H by a repair reaction on the frayed ends and gaps with DNA polymerase. The representation of single copy Drosophila DNA sequences in the library was determined by comparing the kinetics and extent of the reassociation reaction of this tracer with total nuclear DNA and with library DNA as drivers. From the reassociation curves (Fig. 7), we see that the observed extents of reaction are 86 $\pm$ 2% and 83 $\pm$ 2%, respectively, indicating that the library contains 96 $\pm$ 5% of the single copy sequences of Drosophila. The observed $cot_{1/2}$'s of genome and library DNA are 270 and 280 mole sec liter^{-1}, respectively. Thus, the bulk of Drosophila single copy DNA is present in the library in the same relative proportion as in the genome.

Fig. 7

Unfortunately, the accuracy of the reassociation kinetics method for determining representation is limited. It would be particularly important to know whether the library represents 93%, 98% or 99.5% (for example) of

Drosophila sequences. There is no straightforward experimental method with sufficient accuracy for making this determination. As stated above and as shown in Fig. 6, a segment completely overlapping that of pCIT12 in the region 42A could not be discovered in the partial R1 library but was found in the random shear library prepared by J. Lauer (1). (Note, however, that all of the sequences of pCIT12 were found in non-overlapping inserts in the Charon 4-partial R1 library.) The absence of this segment in the partial R1 library probably indicates that one or several of the 5 closely clustered EcoR1 sites in the middle of pCIT12 (Fig. 5) is particularly sensitive to cleavage and is almost always cut even under condition of partial digestion. Nevertheless, in our hands, the partial R1 library has proven to be useful for isolation of a number of single copy genes.

Different recombinant inserts probably affect the burst size and replication time of the bacteriophage differently. Faster growing phage will tend to displace slower growing ones when the library is amplified, and thus the representation of the library will be decreased. It is important to determine conditions for amplifying the library without loss of representation. The relatively large amount of library DNA needed as a driver for the experiment of Fig. 7 was prepared by two cycles of growth of phage from the primary library on a bacterial lawn on nutrient plates, thus minimizing (but not eliminating) competition between fast and slow growers. However, as shown in Fig. 7, when the primary library is amplified by two cycles of growth in liquid culture, the resulting samples contain only about 50% of the single copy sequences present in genomic DNA. Thus amplification by growth as plate stocks is superior to growth in liquid culture for preserving sequence representation.

The theoretical problem of the representation of sequences in a partial digest library has been studied by B. Seed (6). An exact mathematical expression for the fraction of sequences totally missing from the library has been derived. However, this expression is too complex for a satisfactory numerical approximation. Nevertheless, the analysis shows clearly that the longer the insert, the wider the window (the difference between maximum and minimum insert lengths) and the greater the frequency of restriction endonuclease sites, the more representative the library, provided that the conditions for partial digestion are properly chosen.

Acknowledgments

We are indebted to Drs. Dieter Söll and Peter Wensink for several personal communications. The purified tRNAs used were a generous gift from the laboratory of Dr. Dieter Söll. The contribution of Dr. Jay Hirsh in the isolation of the actin gene clone is noted in the text. This research has been supported by grants GM 10991 and GM 20927 from the United States Public Health Service.

References

1. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K., Efstratiadis, A. (1978) Cell, 15, 687-701.
2. Blattner, F.R., Williams, B.G., Blechi, A.E., Thompson, K.D., Faber, H.E., Furlong, L.A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L., Smithies, O. (1977) Science, 1976, 161-169.
3. Sternberg, N., Tiemeier, D., Enquist, L. (1977) Gene, 1, 255-280.
4. Hohn, B., Murray, K. (1977) Proc. Natl. Acad. Sci. USA, 74, 3259-3263.
5. Benton, W.D., Davis, R.W. (1977) Science, 1976, 180-182.
6. Seed, B. (1979). Ph.D. Thesis, California Institute of Technology.
7. Vandekerckhove, J., Weber, K. (1978) J. Mol. Biol., 126, 783-802.
8. Kindle, K.L., Firtel, R.A. (1978) Cell, 15, 763-778.
9. Pollard, T.D., Weihing, R.R. (1974) CRC Crit. Rev. Biochem., 2, 1-65.
10. Kaback, D.B., Angerer, L.M., Davidson, N. (1979) Nuc. Acid. Res., 6, 2499-2517.
11. Southern, E.M. (1975) J. Mol. Biol., 98, 503-517.
12. Taylor, J.M., Illmensee, R., Summers, S. (1976) Biochem. Biophys. Acta, 442, 324-330.
13. Gall, J.G., Pardue, M.L. (1971) Methods in Enzymology, 21D, 470-480.
14. Hunter, T., Garrells, J.I. (1977) Cell, 12, 767-781.
15. Storti, R.V., Horovitch, S.J., Scott, M.P., Rich, A., Pardue, M.L. (1978) Cell, 3, 589-598.
16. Fyrberg, E.A., Donady, J.J. (1979) Devel. Biol., 68, 487-502.
17. Ritossa, F.M., Atwood, K.C., Spiegelman, S. (1966) Genetics, 54, 663-
18. Tartof, K.D., Perry, R.P. (1970) J. Mol. Biol., 98, 503-
19. Weber, L., Berger, E. (1976) Biochem., 15, 5511-

20. White, B.N., Tener, G.M., Holden, J., Suzuki, D.T. (1973). Dev. Biol., 33, 185-195.

21. Steffensen, D.M., Wimber, D.E. (1971) Genetics, 69, 163-178.

22. Elder, R., Szabo, P., Uhlenbeck, O. (1979) In Transfer RNA (Söll, D., Abelson, J., Schimmel, P. eds.) Cold Spring Harbor Laboratory, in press.

23. Tener, G.M., Hayashi, S., Dunn, R., Delaney, A., Gillam, I.C., Grigliatti, T.A. Kaufman, T.C., Suzuki, D.T. (1979) In Transfer RNA (Söll, D., Abelson, J., Schimmel, P., eds) Cold Spring Harbor Laboratory, New York, in press.

24. Dunn, R., Hayashi, S., Gillam, I.C., Delaney, A.D., Tener, G.M., Grigliatti, T.A., Kaufman, T.C., Suzuki, D.T. (1979) J. Mol. Biol., in press.

25. Kubli, E., Schmidt, T. (1978) Nuc. Acid. Res., 5, 1465-1478.

26. Schmidt, T., Egg., A.H., Kubli, E. (1979) Molec. Gen. Genet., in press.

27. Dunn, R., Delaney, A.D., Gillam, I.C., Hayashi, S., Tener, G.M., Grigliatti, T., Misra, V., Spurr, M.G., Taylor, D.M., Miller, R.C., Jr. (1979) Gene, in press.

28. Yen, P.H., Sodja, A., Cohen, M., Conrad, S.E., Wu, M., Davidson, N., Ilgen, C. (1977) Cell, 11, 763-777.

29. Schmidt, O., Mao, J.-I., Silverman, S., Hovemann, B., Söll, D. (1978) Proc. Nat. Acad. Sci. USA, 75, 4819-4824.

30. Hoevemann, B., Schmidt, O., Yamada, H., Silverman, S., Mao, J.-I., DeFranco, D., Söll, D. (1979) In Transfer RNA (Söll, D., Abelson, J., Schimmel, P., eds.) Cold Spring Harbor Laboratory, New York, in press.

31. England, T.E., Uhlenbeck, O.C. (1978) Nature, 280-281.

32. Galau, G.A., Klein, W.H., Davis, M.M., Wold, B.J., Britten, R.J., Davidson, E.H. (1976) Cell, 7, 487-505.

Table 1

Drosophila 4S RNA Sites[†a]

X	2L	2R	3L	3R	4
3D	22DE	41CD	61D Met3	84A Met2	none
3F	23EF Ser-4,7	*42A Asn Arg2, Lys2	*62A Lys2 Glu4	84D Val3b	
5F6A	24DE	42E Lys2	63A	84F Arg2	
11A	*27D	*43A	63B	87BC	
*12E Ser-4,7	28O	44EF	64DE Val3a	87F88A	
	28F29A	46A Met3	66B	89BC Val14	
	29D	48AB Met2	67B	90C Val3b	
	29E Asp2	48CD	69F70A	90DE	
	*34A	49AB	70BC Val4	*91C	
	35AB	49F50A	70DE	92AB Val3b	
		*50BC Lys2	72F*73A Met2	93A	
		52F53A Glu4	79F	95F96A	
		54A		97CD	
		54D		99EF	
		55EF			
		56D Val4			
		*56F Glu4			
		57D			
		58AB			
		60E			

[†a]The locus assignments in heavy type were made by Elder et al. (22) using total labeled tRNA. Sites that were heavily labeled are underlined. Entries to the side of the main columns of the table are assignments made using purified tRNA's (21,22,23,24,25,26,27). Asterisks indicate assignments for recombinant Charon plaques that give strong tRNA signals, isolated in our laboratory. Entries in italics are in situ assignments that may be different from those of Elder et al. (22).

Legends

Figure 1. Construction of an EcoR1 partial digest eukaryote library, with Charon 4 as a vector, as described in the text.

Figure 2. Restriction endonuclease map of the Drosophila insert of λDmA2 bearing an actin gene. In the lower figure, the solid bars indicate the regions present in Drosophila actin mRNA.

Figure 3. Electron micrographs of R-loops of Drosophila actin mRNA hybridized to λDmA2 by the method of Kaback et al. (10). B is an interpreted tracing of the R-loop region in micrograph A. D is an enlargement of the important region in C, and E is an interpreted tracing.

Figure 4. Restriction endonuclease digests and Southern blots of the 8.7 kb EcoR1 fragment of λDmA2 (Fig. 2) subcloned in pBR322. The plasmid DNA was digested with restriction endonucleases as indicated, and subjected to gel electrophoresis on a 1.0% agarose slab gel. Panel A shows the ethidium bromide staining pattern of the gel, and thus displays all of the restriction fragments from the map of Fig. 2 (plus those of the vector). For panels B and C, identical gels were run in parallel and the DNA blotted on to nitrocellulose paper. In panel B, the ^{32}P labeled cDNA probe was prepared by reverse transcription of total poly(A^+) larval RNA with a calf thymus digested primer (12). In panel C, the same RNA was reverse transcribed using an oligo (dT) primer. In either case, the

only component of the RNA which gives detectable hybridization with λDmA2 DNA is the actin sequences. The probe for panel B is uniformly representative of the entire actin mRNA; that for panel C is biased towards the first few hundred nucleotides on the 3' side of the message. With the representative probe, the 0.8 kb Sal fragment hybridizes more strongly than the 1.1 kb fragment. With the 3' biased probe, the 1.1 kb fragment gives the stronger signal. These observations lead to the 5' to 3' orientation of the actin gene shown in Figure 2. The relative intensities of hybridization to the 1.8 and 1.6 kb Hind III fragments confirm the interpretation.

Figure 5. Map of the pCIT12 plasmid showing the locations of the restriction sites, the inverted repeats and the tRNA genes. Only the Drosophila region is presented. The bars for the genes for tRNA-1, -2, and -4 represent the standard deviation of electron microscope measurements. The tRNA-3 gene was identified and mapped only by Southern blotting with restriction fragments. Restriction fragments with tRNA genes are indicated as solid bars. This part of the figure is adapted from Yen et al. (28). The mapping data for the 9 tRNA genes shown at the top of the graph are from the sequencing studies reported by Hovemann, et. al. (30). The inverted repeat stem c is a $tRNA^{Ile}$ gene, the stems a, a' and c' are for $tRNA^{Lys}$ genes, whereas b_2, b_1, and b' are $tRNA^{Asn}$ genes. Directions of transcription as determined by sequencing are shown.

map of DM

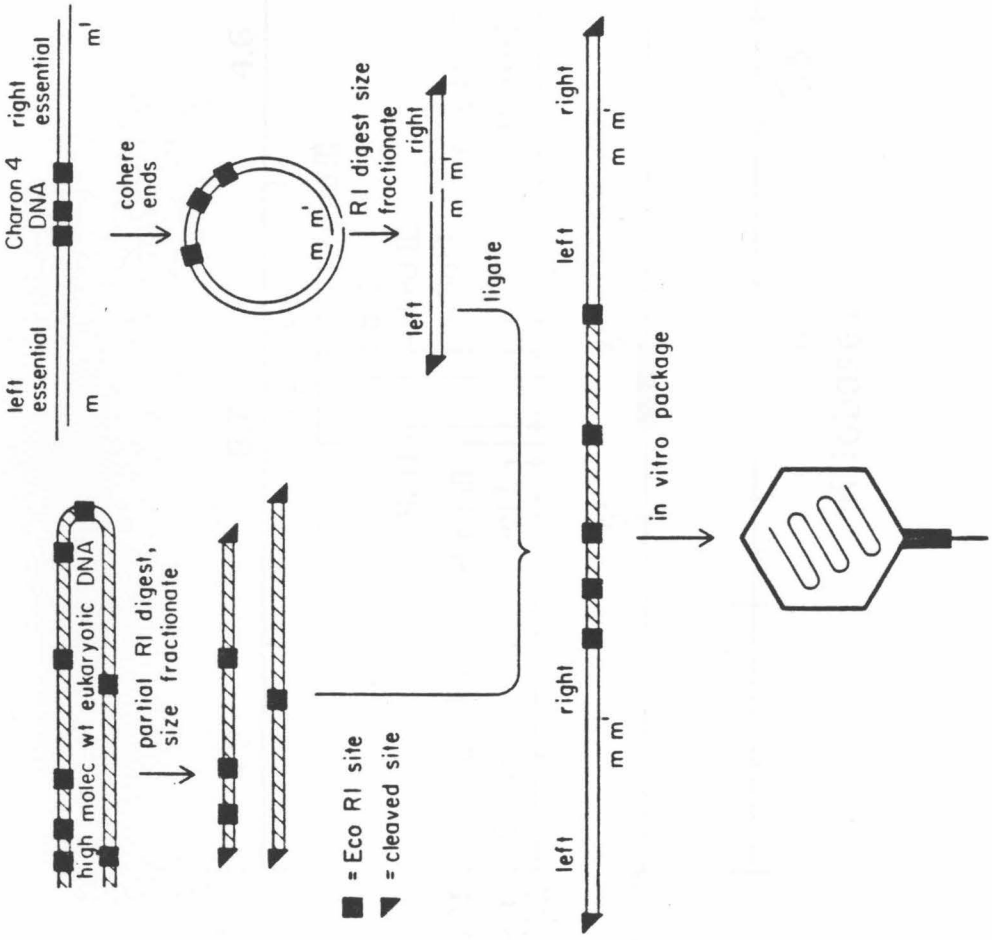
Figure 6. DM inserts in Charon 4 phage covering the region 42A. The scale is in kilobases (kb). ↓ denote EcoR1 cleavage sites. Capital

letters denote EcoR1 restriction fragments referred to in the text. Note that fragment A of pCIT12 contains the Col E1 vector as well as segments on both ends of the insert. Positions of tRNA genes were determined by blotting experiments with labeled total or purified tRNA probes as described in the text. The bacteriophage with numbers greater than 100 were obtained by screening the random shear library prepared by J. Lauer (1); all other inserts were obtained from the partial R1 library.

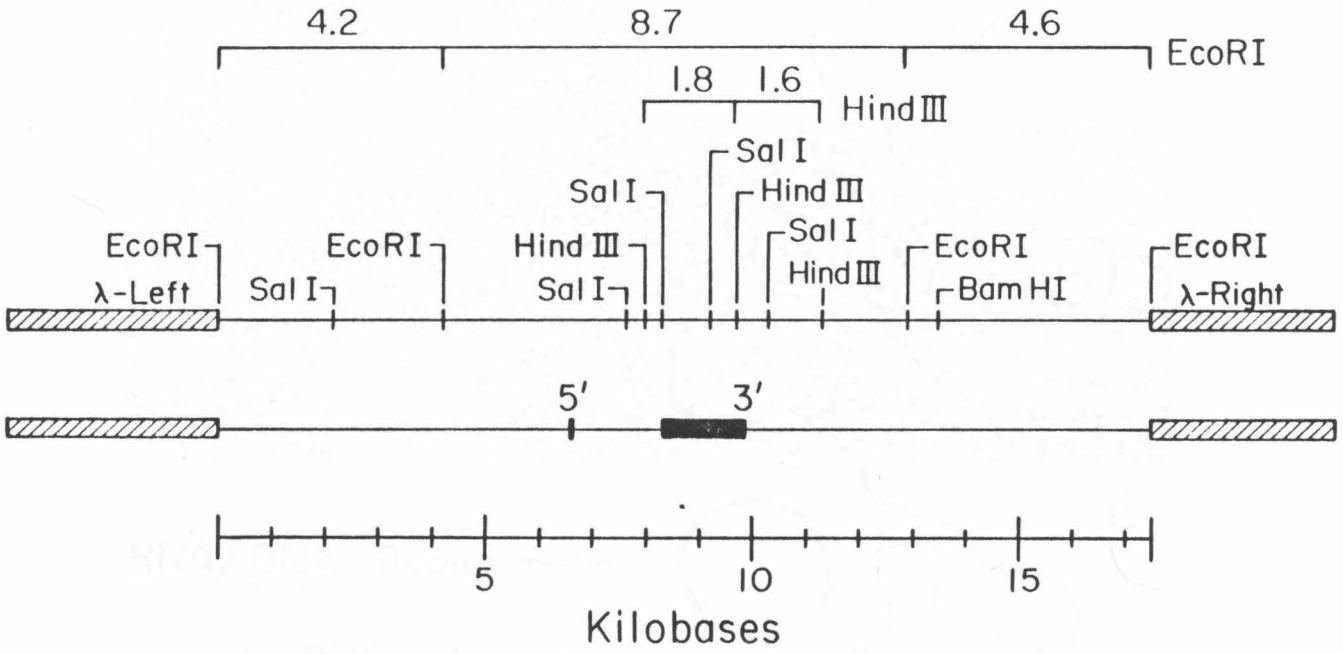
Figure 7. Determination of the complexity of the Drosophila DNA sequences in the partial R1 library. ^3H labeled single copy Drosophila DNA was prepared by the procedure of Galau et al. (32) as follows. Sheared Drosophila pupal nuclear DNA at a concentration of 0.37 mg/ml was incubated in 0.41 M phosphate equivalent buffer (0.41 M PB, Na^+ = 0.62 M (32)) at 67° for 9.5 hr (Cot = 200) and passed over a HAP column in 0.12 M PB at 60°. The single stranded fraction was isolated, reacted to a Cot of 100 and again passed over HAP. The resulting single stranded material was incubated to a Cot of 5000; the double stranded fraction was isolated and labeled by gap translation to a specific activity of approximately 10^6 cpm/μg.

Time points were taken by incubating driver DNA at concentrations of 0.1 to 10 mg/ml in 0.41 M PB at 67° for times up to 50 hrs. Single and double stranded fractions were assayed by HAP chromatography. The horizontal axis on the figure is equivalent Cot (in 0.12 M PB). We use a salt correction from 0.41 M PB of 5.0; we assume that the average mass fraction of Dm DNA in total library DNA is 0.35.

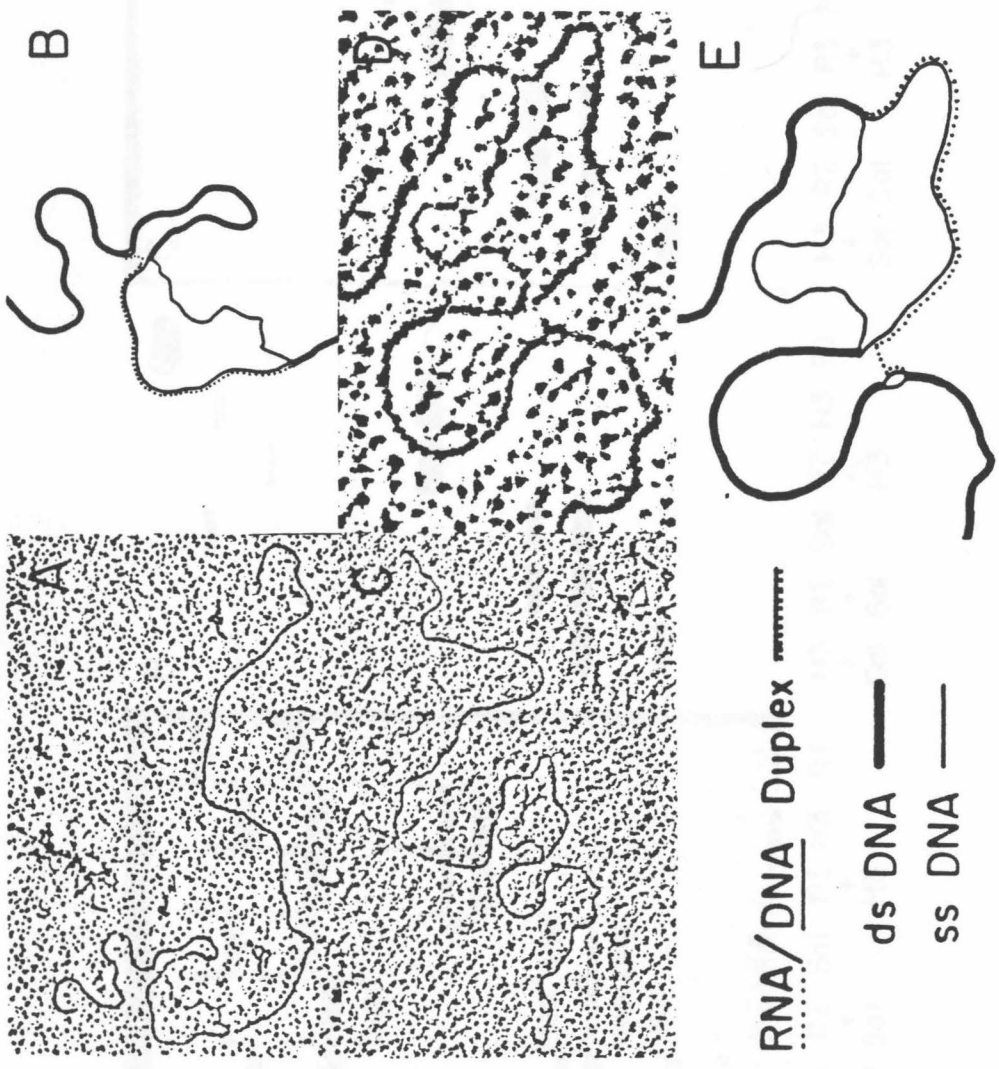
Amplified plate stocks of library phage were prepared from ten 150 mm plates, each with 10^4 phage from the primary library, plated on to a bacterial lawn of KH 802. From the resulting phage, 10^6 were plated onto 10 new plates from which 900 µg of DNA was isolated. For the liquid culture experiments 10^5 phage were adsorbed onto 10^8 bacteria in 1 ml, then diluted and grown in 10 ml, following the general PDS procedure (1). The yield was approximately 1.3×10^{11} phage. 4×10^7 phage were then reinoculated into 4×10^{10} bacteria and grown in one litre of culture by the previous methods, yielding approximately 1000 µg of DNA.


left essential
Charon 4 DNA
right essential
m
m'
cohere ends
m m'
RI digest size fractionate
left
right
m m'
ligate
high molec wt eukaryotic DNA
partial RI digest, size fractionate
■ = Eco RI site
▼ = cleaved site
left
right
m m'
in vitro package

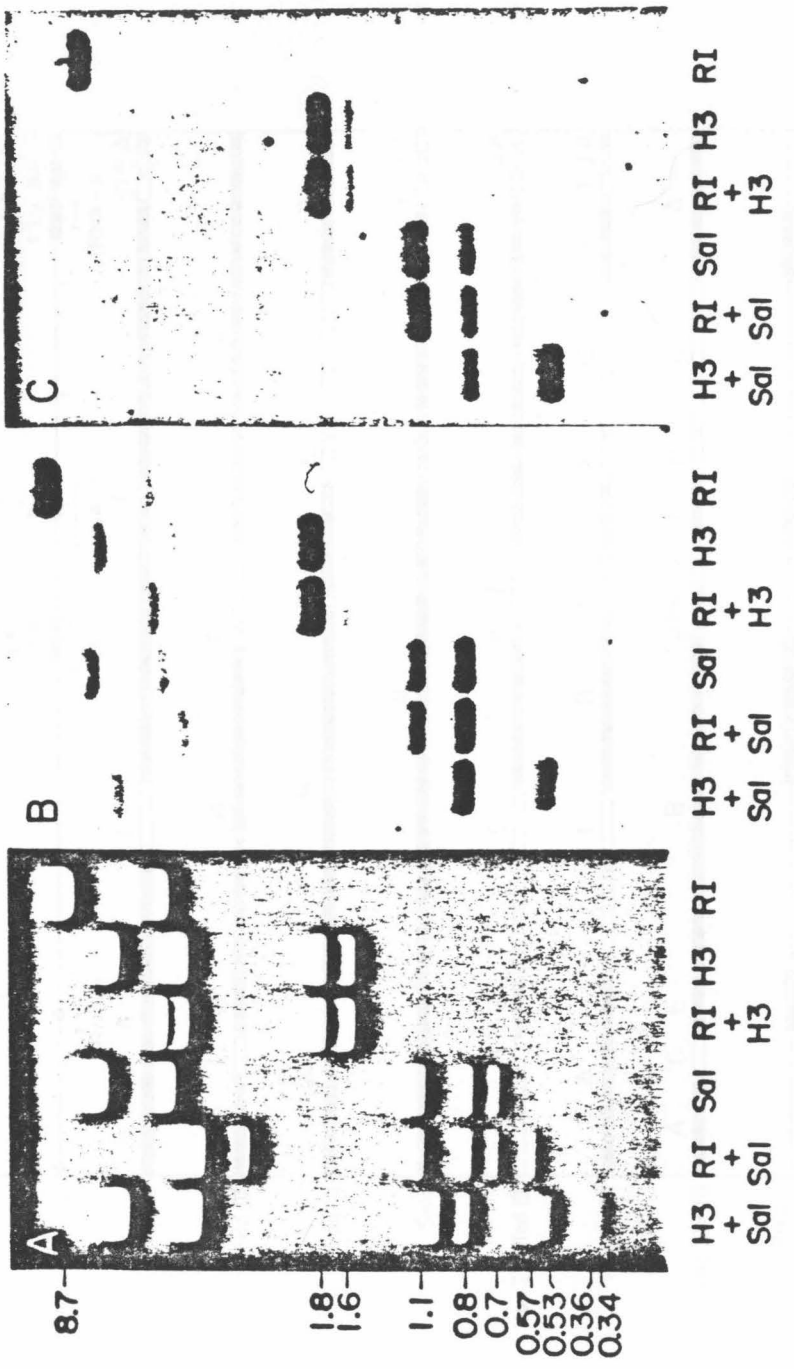
F1


4.2
8.7
4.6
EcoRI
1.8
1.6
Hind III
Sal I
Sal I
Hind III
Sal I
Hind III
EcoRI
EcoRI
EcoRI
EcoRI
Hind III
Sal I
Sal I
Hind III
Bam HI
λ-Left
λ-Right
5'
3'
5
10
15
Kilobases

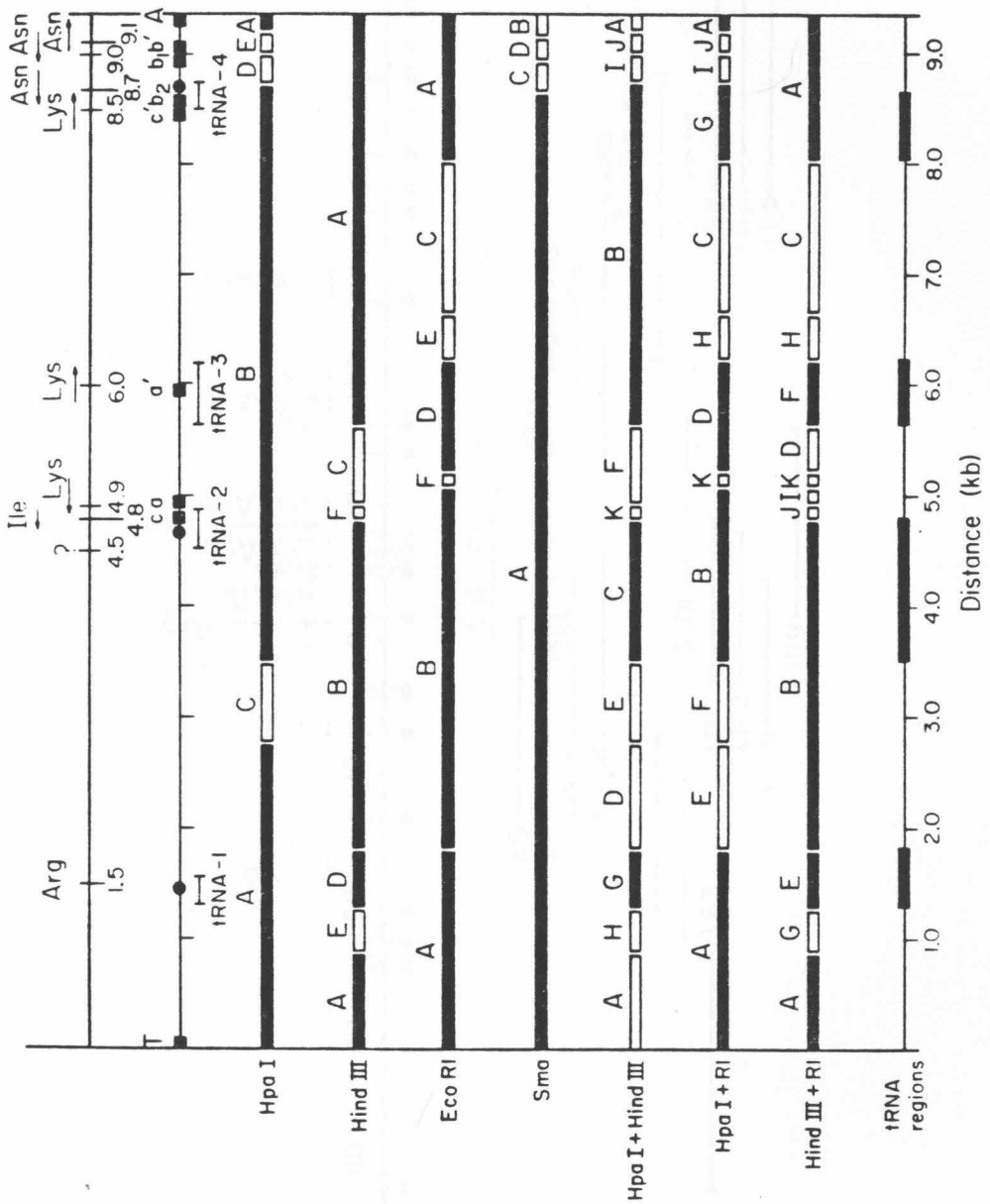
F2



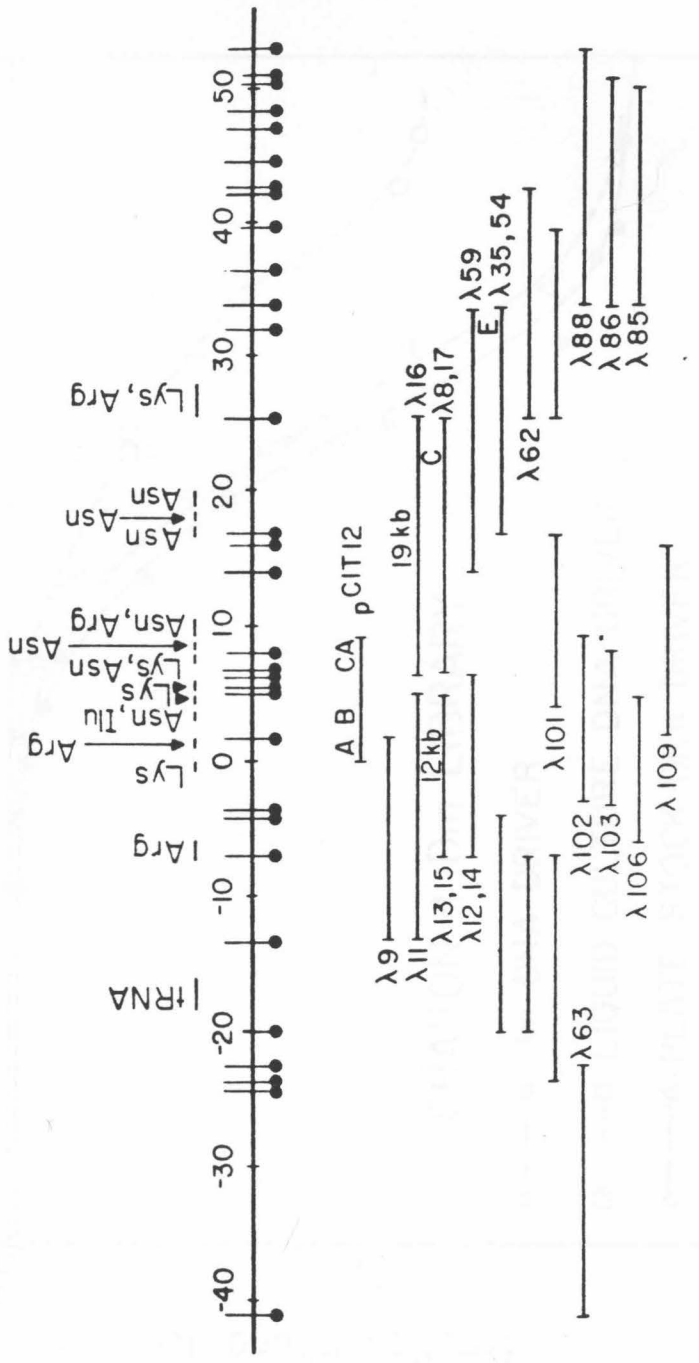
F3


A
B
C
8.7
1.8
1.6
1.1
0.8
0.7
0.57
0.53
0.36
0.34
H3 RI Sal RI H3 RI
+ + +
Sal Sal H3
H3 RI Sal RI H3 RI
+ + +
Sal Sal H3
H3 RI Sal RI H3 RI
+ + +
Sal Sal H3

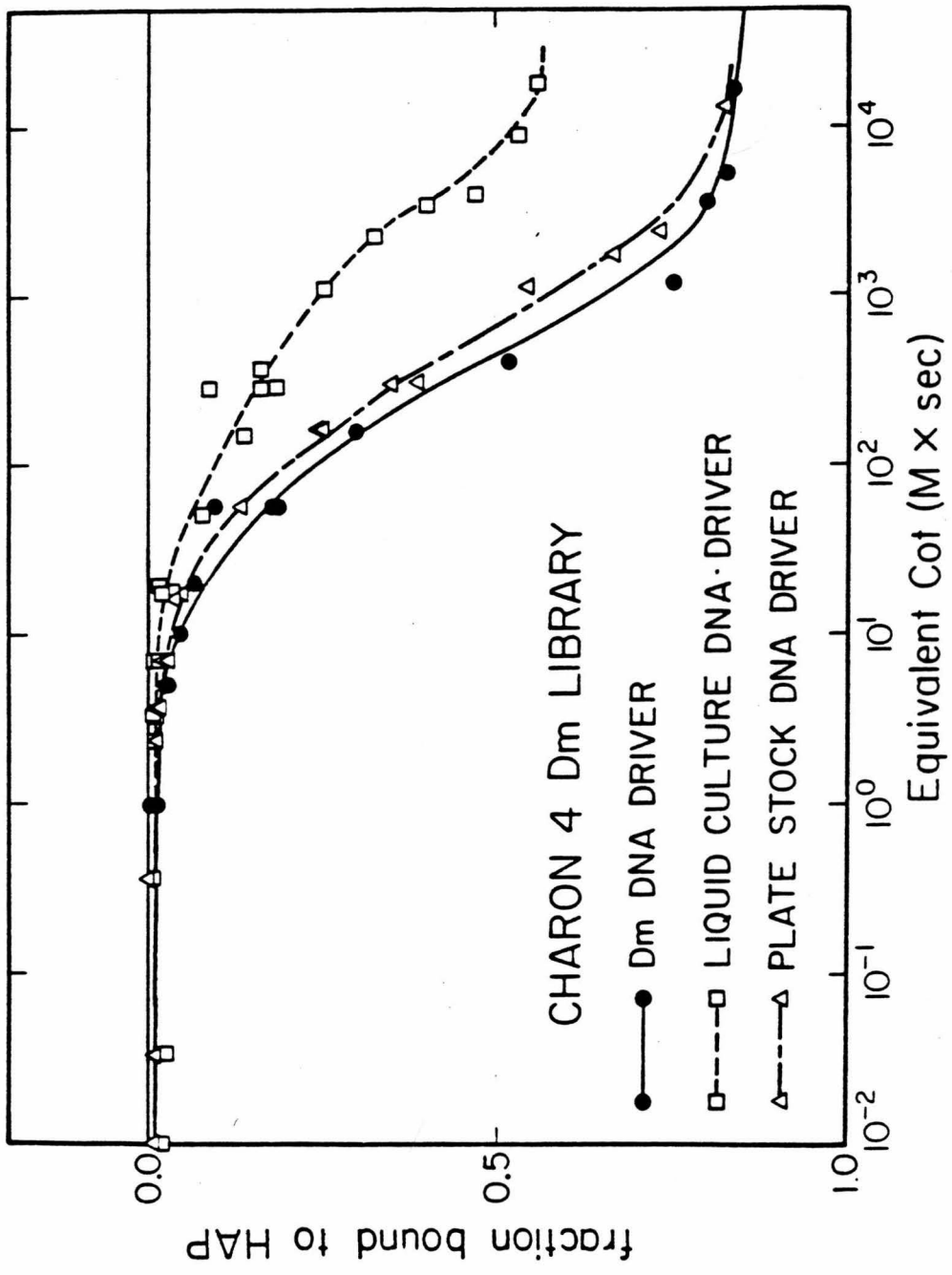
F4


Asn Asn
Lys Asn
Ile
Lys
Lys
Arg
tRNA-1
tRNA-2
tRNA-3
tRNA-4
Hpa I
Hind III
Eco RI
Sma
Hpa I + Hind III
Hpa I + RI
Hind III + RI
tRNA regions
Distance (kb)

F5


-40
-30
-20
-10
0
10
20
30
40
50
tRNA
Arg
Lys
Asn, Ile
Arg
Lys
Lys, Asn
Asn
Asn, Arg
Asn
Asn
Lys, Arg
A B CA pCIT12
19 kb
12kb
λ9
λ11
λ13,15
λ12,14
λ16
λ8,17
C
λ59
λ35,54
E
λ62
λ88
λ86
λ85
λ101
λ102
λ103
λ106
λ109
λ63

F6


fraction bound to HAP
0.0
0.5
1.0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
Equivalent Cot (M x sec)
CHARON 4 Dm LIBRARY
Dm DNA DRIVER
LIQUID CULTURE DNA·DRIVER
PLATE STOCK DNA DRIVER

F7