SERINE tRNAs AND THEIR GENES IN *DROSOPHILA MELANOGASTER*

by

DAVID LAMAR CRIBBS
B.A., The University of Colorado, 1975
M.Sc., The University of British Columbia, 1979

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies
Department of Biochemistry
Faculty of Medicine

We accept this thesis as conforming

to the required standard

The University of British Columbia
January 1982

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Biochemistry

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date Feb. 11, 1982

SUMMARY

Serine tRNAs and their genes in Drosophila melanogaster were characterized. The nucleotide sequences of $tRNA_4^{Ser}$ (codon UCG), $tRNA_7^{Ser}$ (UCA, UCC, UCU), and $tRNA_{2b}^{Ser}$ (AGC, AGU) were determined. Also, the nucleotide sequences of four $tRNA^{Ser}$ genes from the X chromosome isolated in recombinant plasmids were determined.

Transfer RNA_4^{Ser} and $tRNA_7^{Ser}$ differ at only three out of 85 positions, including the "wobble" nucleotide of the anticodon. However, $tRNA_{2b}^{Ser}$ is only 72% homologous with $tRNA_4^{Ser}$ (62 out of 85 positions, not counting differences in modification). Major regions of sequence homology (> 5 consecutive positions) are found only in the D arm (21 consecutive positions) and in the TψC arm (11 consecutive positions).

Transfer RNA_4^{Ser} and $tRNA_7^{Ser}$ are indistinguishable by RNA-DNA hybridization. Both hybridize to the same sites on polytene chromosomes in situ, including the major site at 12DE on the X chromosome, and 23E on chromosome 2 (Hayashi et al. (1980). Chromosoma 76, 65-84.) No other purified tRNA than $tRNAs_{4,7}^{Ser}$ has been shown to hybridize to the X chromosome in Drosophila. Therefore, several X-derived recombinant plasmids hybridizing $tRNA_{4,7}^{Ser}$ (pDt 16, pDt 17R, pDt 27R, and pDt 73; Dunn et al. (1979). Gene 7, 197-215.) were analyzed. Based on the results of Southern blotting experiments, there appear to be eight $tRNA^{Ser}$ genes on the four plasmids (one each on pDt 17R and pDt 73; two on pDt 16; and four on pDt 27R). Thus, the 12DE region contains at least eight $tRNA^{Ser}$ genes.

Restriction mapping and DNA sequence analysis were performed with pDt 16, pDt 17R, and pDt 73. Based on the tRNA sequences, which differ at three positions, the presumptive DNA sequences encoding $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ can be represented as 444 or 777 genes. DNA sequence analysis gave surprising results in this respect. Analysis of four $tRNA^{Ser}$ genes on the three plasmids identified two 777 genes matching $tRNA_7^{Ser}$ plus "hybrid" 774 and 474 sequences. Further, pDt 16 contains both a 777 and a 774 gene as a direct repeat 400 base pairs apart. Since a 444 gene corresponding to $tRNA_4^{Ser}$ must exist, there are at least four different types of closely related serine tRNA genes in the D. melanogaster genome. This observation may have implications concerning the evolution and maintenance of reiterated tRNA genes in eukaryotes.

In addition to studies on serine tRNAs and their genes, the nucleotide sequences of Drosophila $tRNA_5^{Lys}$ and of a $tRNA^{Arg}$ were determined.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I wish to thank Gordon Tener for his help, encouragement, and consideration as my graduate supervisor. I also wish to thank Ian Gillam for his many contributions to my research. Finally, I am in the debt of a large number of people in our laboratory and in the several groups collaborating with us for discussions, advice, gifts of materials, and enthusiasm.

DEDICATION

This thesis is dedicated to the memory of my father, and to my mother.

ABBREVIATIONS

A number of abbreviations are used routinely throughout the text. For convenience, these are compiled here.

Nucleosides and nucleotides:

N	nucleoside (A,G,C,U, or modified nucleosides)
dN	deoxynucleoside (dA,dG,dC,dT)
NMP	nucleoside 5'-phosphate
dNMP	deoxynucleoside 5'-phosphate
NDP	nucleoside 5'-diphosphate
dNDP	deoxynucleoside 5'-diphosphate
NTP	nucleoside 5'-triphosphate
dNTP	deoxynucleoside 5'-triphosphate

A	adenosine
G	guanosine
C	cytidine
U	uridine
dA	deoxyadenosine
dG	deoxyguanosine
dC	deoxycytidine
dT	deoxythymidine (thymidine)

Modified nucleosides are abbreviated as recommended in ref. 12.

p	phosphoryl group
*p	^{32}P-labelled phosphoryl group
pN	nucleoside 5'-phosphate
Np	nucleoside 3'-phosphate
Np (2')	nucleoside 2'-phosphate
pNp	nucleoside 5', 3'-bisphosphate
*pN	nucleoside [5'-^{32}P] phosphate
Np*	nucleoside [3'-^{32}P] phosphate
*pNp	nucleoside 5', 3'-bis [5'-^{32}P] phosphate

Other:

XC	xylene cyanol FF
BB	bromophenol blue
TEMED	N,N,N',N'-tetramethylethylenediamine
SSC	0.15 M NaCl, 0.015 M sodium citrate, pH 7
2 x SSC	0.30M NaCl, 0.030 M Sodium citrate, pH 7
NH_4OAc	ammonium acetate
b.p.	base pairs
kb	kilobase pairs

Chapter I

INTRODUCTION

A. Preface

This thesis deals with a study on serine tRNAs and serine tRNA genes in Drosophila melanogaster. It presents data on the nucleotide sequences of several Drosophila tRNAs, including $tRNAs^{Ser}$ reading the six serine codons, and on the structure and organization of $tRNA^{Ser}$ genes.

The thesis is divided into five chapters. They are presented in the following way. (i) The introductory chapter (Chapter I) is intended to provide the reader with an overview of the characteristics of tRNA and of tRNA genes. The last section of the introduction deals specifically with Drosophila tRNAs and tRNA genes. In it I emphasize advances of the last several years, largely made possible by recombinant DNA techniques. (ii) Chapter II contains a description of methods used in the present work. (iii) Nucleotide sequence analyses of three serine isoacceptor tRNAs from Drosophila are presented in Chapter III. (iv) The fourth chapter deals with the structure and organization of serine tRNA genes in Drosophila. The analysis combines data from in situ hybridization experiments by other workers in this laboratory with restriction enzyme mapping and nucleotide sequence analysis of recombinant plasmids hybridizing $tRNAs^{Ser}$. In particular, structure and organization of $tRNA^{Ser}$ genes obtained from the 12DE region on

the X chromosome are considered. (v) Under some conditions, $tRNA_5^{Lys}$ hybridizes to the 12DE region of polytene X chromosomes *in situ*. As part of a comprehensive characterization of this region at the molecular level, the nucleotide sequence of $tRNA_5^{Lys}$ was determined. A partial sequence of a $tRNA^{Arg}$ contaminating the $tRNA_5^{Lys}$ preparation is also presented.

An introduction relevant to eukaryotic serine tRNAs and their genes is presented in Chapter III, but applies also to Chapter IV. A separate introduction concerning *Drosophila* lysine tRNAs and their genes is found in Chapter V. Chapters III, IV, and V each present results and discussion pertaining to the research problem of that chapter. Discussion of the major results of the thesis research is found in Chapter IV, "Structure and Organization of Serine tRNA Genes in *Drosophila*."

B. Structure of Transfer RNAs

Transfer RNA (tRNA) is a class of molecules functioning as carriers of activated amino acids in protein biosynthesis, a fundamental process in any living cell. It is an ancient class of molecules, probably having existed in some form since the first living cell (1). Transfer RNA molecules are remarkable for the large number of cell components with which they interact. These include processing and modifying enzymes, nucleotidyl transferase, aminoacyl-tRNA synthetases, initiation factors, elongation factors, and ribosomes.

Transfer RNAs have a very central role in cellular metabolism. They function to insert amino acids into nascent polypeptide chains in response to DNA-coded mRNA. Thus, tRNAs are responsible for transforming

the linear, sequence-encoded information of DNA into protein structures responsible for dynamic expression of the genetic information stored in the genome. For these reasons, tRNA structure and function have been the subjects of extensive investigation.

B.1. Structure of tRNA: The "Cloverleaf"

The existence of transfer RNA, predicted earlier on theoretical grounds by Crick, was demonstrated by Hoaglund and Zamecnik in 1957 (2). The first nucleotide sequence of a purified tRNA was completed by Holley and his co-workers in 1965 (3). This sequence, of yeast alanine tRNA, can be drawn in several possible secondary structures. One of these, the "cloverleaf" structure, was also a possible form for several tRNAs whose sequences were completed shortly after (4,5,6). The nucleotide sequence of yeast phenylalanine tRNA is shown in its "cloverleaf" structure in Figure 1.

The sequences of many tRNAs are now known (7). All cytoplasmic tRNAs can be represented by a standard "cloverleaf", with invariant or semi-invariant nucleotides in characteristic positions (Figure 2). Reviews on this subject are available (8,9). Transfer RNAs generally contain 75-90 nucleotides (7). The standard structure consists of an acceptor arm, the dihydrouridine (D) arm containing the D loop (Loop I), the anticodon arm containing the anticodon loop (Loop II), the extra or variable arm (Loop III), and the TψC arm and loop (Loop IV). The acceptor arm consists of a stem containing seven base pairs (usually Watson-Crick base pairs, G-C or A-U) and a single-stranded tail at the 3'-terminus with sequence NCCA.

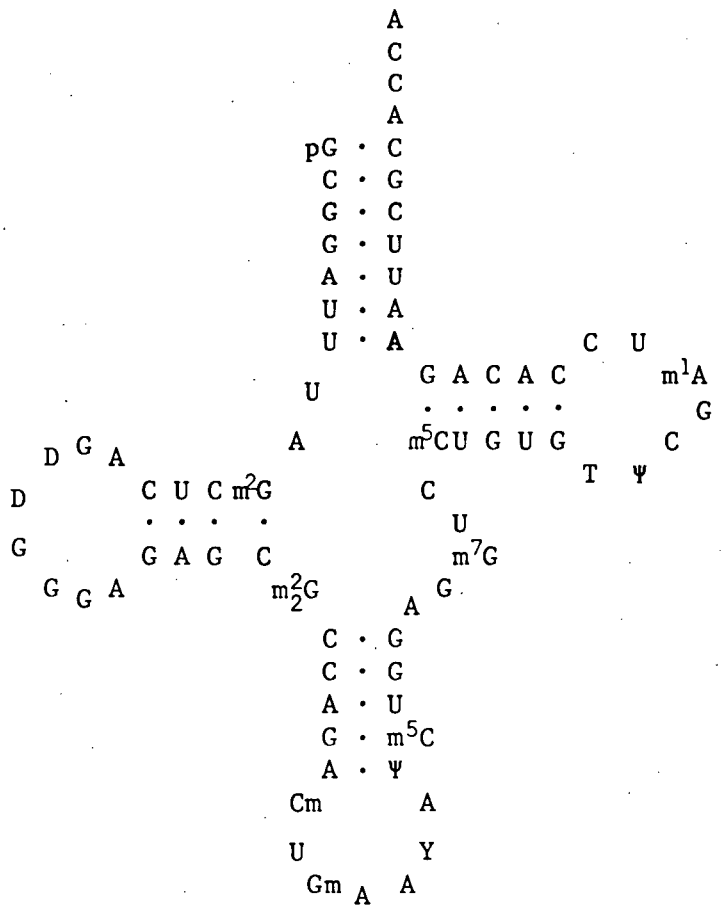


Figure 1 - Cloverleaf structure of yeast $tRNA^{Phe}$.

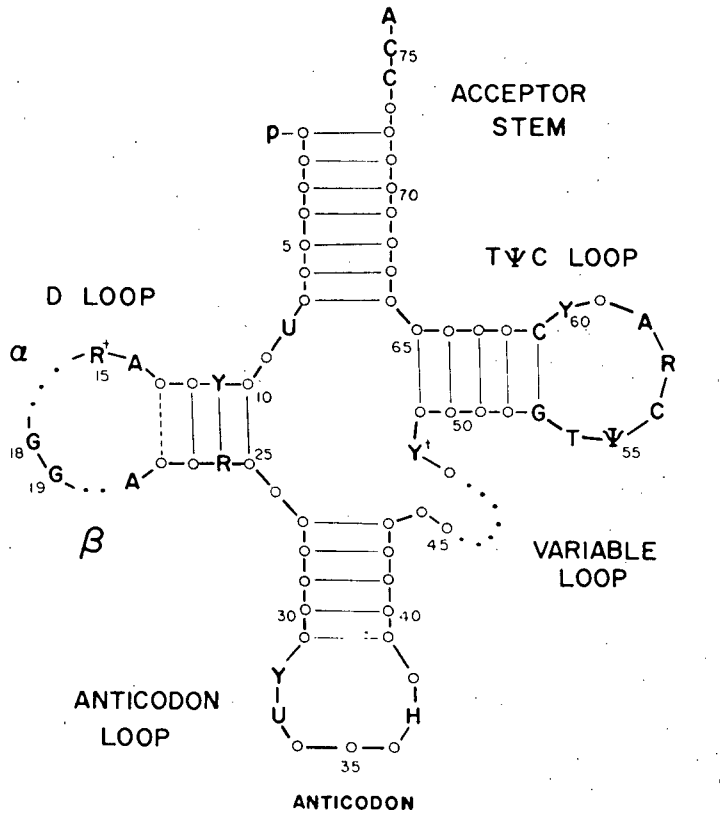


Figure 2 - Generalized cloverleaf structure of tRNA.

(An important exception is the class of prokaryotic initiator methionine tRNAs, in which the 5'-terminal base and the base opposite it do not pair. This feature is apparently necessary for correct recognition by specific initiation factors. Also, histidine tRNAs from both prokaryotes and eukaryotes have an extra nucleotide at the 5'-terminus, which may or may not base pair with the opposite nucleotide.) Aminoacylation of the tRNA occurs at the 3'-terminal adenosine residue. The D arm contains a stem of three or four base pairs and a loop. The D loop is of variable length, usually 8 to 10 nucleotides, with variable α and β regions flanking the standard GG sequence at positions 17 and 18. The D loop normally contains at least one dihydrouridine residue. The anticodon arm comprises a stem of five base pairs and a loop of seven nucleotides containing the triplet anticodon. The extra or variable arm is four or five nucleotides long in about 80% of cytoplasmic tRNA sequences, but 13-21 nucleotides long in the remaining 20%. (The extra arm and the α and β regions of the D loop are the only sequences of variable length in the standard "cloverleaf".) The TψC arm is made up of a five base-pair stem and a seven nucleotide loop. The first (5') nucleotide of the loop is usually ribothymidine (T).

B.2. Modified nucleosides of tRNA

Transfer RNAs contain many different types of modified nucleosides. These have been discussed in a number of reviews (10,11,12,13,14). Up to 25% of the nucleosides in a tRNA may be modified (7). Each modification, introduced post-transcriptionally by specific tRNA-modifying enzymes (15), occurs in one or a few characteristic positions in the tRNA "cloverleaf"

(7). The modified nucleosides of tRNA have been divided into three classes: (i) those found in the first (5') or "wobble" position of the anticodon; (ii) those found adjacent to the 3' end of the anticodon; and (c) those found elsewhere (12,13,14).

B.2.a. Modified nucleosides in the first position of the anticodon

The genetic code consists of 64 triplet codons, grouped in fours as XYN (e.g. CGN, where X = C, Y = G, and N = A,G,C, or U). Of the 64 codons, 61 code for an amino acid ("sense" codons); the other three signal termination of translation ("nonsense" codons). In a number of codon groups, all four specify the same amino acid. Crick's "wobble" hypothesis sets out rules for pairing between the anticodons of tRNAs and codons in mRNA (16). It allows non-standard base-pairs between the first nucleoside of the anticodon and the third nucleoside of the codon, together with standard A-U or G-C base-pairs between tRNA and the first two codon letters. By allowing non-standard base-pairs, the degenerate set of 61 sense codons can be read by considerably less than 61 tRNAs. For example, Crick suggested that tRNAs with G in the "wobble" position would read codons ending in C and U; that tRNAs with inosine (I) there would read codons ending in A, C, or U; and that tRNAs with U in this position would read those ending in A and G. The first and second cases are certainly true. Inosine-containing tRNAs respond to codons ending in A, C, or U in ribosome binding assays. Similarly, tRNAs with G in the first anticodon position read codons ending in U or C (16,17). On the other hand, virtually no cytoplasmic tRNAs have been found with an unmodified U in the "wobble" position (7,14). Mitochondrial tRNAs from Neurospora crassa contain U in this position, however, and presumably pair as predicted (18).

A number of modified nucleosides are found only in the first position of the anticodon. These includes uridine-5-oxyacetic acid (V base), 5-methoxycarbonylmethyl-2-thiouridine (mcm^5s^2U), 5-methylaminomethyl-2-thiouridine (mam^5s^2U), and 7-(4,5-cisdihydroxy-1-cyclopenten-3-ylamino-methyl)-7-deazaguanosine (Q base). Modifications of the "wobble" position alter codon specificity of the tRNAs. Whether specificity is increased or decreased depends on the particular modification. Conversion of uridine to V base allows decoding of codons ending in U as well as A and G. Conversion of adenosine to inosine results in ability to read codons ending in A, C, and U(12,13,14). These two modifications thus result in decreased codon specificity. However, mcm^5s^2U and Q base seem to function to increase codon specificity. Transfer RNAs containing Q base (a derivative of guanine) respond preferentially to codons NAU over NAC (14), while rabbit liver $tRNA^{Lys}$ containing mcm^5s^2U responds to AAA in preference to AAG (19). Yeast $tRNA_3^{Glu}$, which contains mcm^5s^2U, responds only to GAA, not GAG (20). Presumably such alteration of codon specificity enhances the functional efficiency of the tRNA population in protein synthesis.

B.2.b. Modified nucleosides located adjacent to the anticodon

A number of modified purine nucleosides are found adjacent (3') to the anticodon. Modifications range from simple methylation to rather exotic hypermodifications. The type of modification correlates well with codon specificity (13). The hypermodified bases fall into two general classes. (1) Transfer RNAs reading codons that begin with U have a modified base with a bulky hydrophobic substituent. Usually this base is N-6-(Δ^2-

isopentenyl)-adenosine (i^6A) or its 2-methylthio derivative, ms^2i^6A. The former is found in eukaryotes, the latter in prokaryotes. One outstanding exception to this generalization is yeast $tRNA^{Phe}$, which contains the hydrophobic Y base (12,13,14). (2) The set of tRNAs reading codons that begin with A contain a base with a bulky hydrophilic substituent. This base is generally N-[9-(β-D-ribofuranosyl) purin-6-ylcarbamoyl] threonine (t^6A) or one of its derivatives such as mt^6A or ms^2t^6A (12,14).

These hypermodified nucleosides are thought to increase translational accuracy for codons beginning with A or U. Transfer RNAs might be expected to interact relatively weakly with codons ANN or UNN, since standard A-U base pairs can form only two hydrogen bonds compared with three for G-C pairs. How the hypermodified bases function is not known. It has been suggested that these bases may extend codon binding to four bases so far as tRNA is concerned (12). An attractive alternative proposal is that these bases prevent distortion of the anticodon and maintain correct base stacking, particularly at the third anticodon base, thereby helping to maintain high translational accuracy at UNN or ANN codons (21).

Nucleosides found adjacent to the anticodon in tRNAs reading CNN or GNN codons include 1-methylguanosine, 2-methyladenosine, 6-methyladenosine, and 1-methylinosine (7).

B.2.c. Modified nucleosides in other positions of the tRNA molecule

Modified nucleosides located in other positions of the tRNA molecule than the "wobble" position or adjacent to the anticodon may function in binding of tRNA to ribosomes, stabilization of the tRNA conformation,

enhancement of nuclease resistance or of specific recognition by aminoacyl-tRNA synthetase (12). While mutant strains of microorganisms lacking some modifying enzymes may appear to grow normally, these modifications may enhance tRNA function sufficiently to confer a selective advantage. For example, bacteria lacking the enzyme that synthesizes ribothymidine from uridine in tRNA are quickly subverted by wild type strains in mixed cultures (22).

Modified nucleosides found in tRNA at positions outside the anti-codon loop include dihydrouridine in the D loop of most tRNAs, Tψ in positions 54 and 55 of most tRNAs, 1-methyladenosine in position 58 of many eukaryotic tRNAs, and 7-methylguanosine and 5-methylcytosine in the extra arm of many tRNAs (7).

B.3. Three-dimensional configuration of tRNA

The determination to 0.25 nm resolution of a three-dimensional structure for yeast $tRNA^{Phe}$ by X-ray crystallography was reported by two independent groups in 1973-74 (23,24). Because the three-dimensional structure is known, $tRNA^{Phe}$ has been the subject of many studies on structure/function relationships. Of particular significance, results of chemical modification studies on this tRNA in aqueous solution give results consistent with the three dimensional structure of crystalline $tRNA^{Phe}$ (25). The crystal structure deduced is therefore relevant to the structure of the molecule in aqueous solution and to its biological function. Many studies consistent with this assertion have been carried out (reviewed in 8,26). However, a detailed discussion of these is not necessary for our purposes here.

A diagrammatic three dimensional structure for yeast $tRNA^{Phe}$ is presented in Figure 3. The molecule has a compact L-shape, with the anticodon loop accessible to read mRNA at one end and the NCCA of the acceptor arm extended at the other end to donate an amino acid into a polypeptide in response to the correct codon.

The structure of various tRNAs must be similar, since all tRNAs active in protein biosynthesis must interact with the ribosomes. However, tRNAs must be different enough to allow highly specific aminoacylation by cognate synthetases, and for recognition of methionine initiator tRNA by specific initiation factors. The existence of such similarity is supported by the fact that homogeneous crystals can be prepared containing a mixture of tRNAs (27). With respect to differences, three dimensional structures determined by X-ray crystallography for yeast initiator $tRNA^{Met}$ (28) and for a mixture of yeast $tRNA^{Gly}$ species (27), while generally similar to that for yeast $tRNA^{Phe}$, are nonetheless distinctive.

The various modified nucleosides may function by altering the surface of the folded tRNA in specific ways to enhance particular tRNA functions. The substituents of modified nucleosides may increase the surface area of a tRNA substantially. By placing distinctive modifications in positions not involved in general tRNA recognition for protein synthesis, specific recognition of generally similar tRNA molecules might be achieved.

B.4. Purification of tRNAs

A number of different column chromatographic systems are used in the purification of tRNAs (discussed in 8,29). These include chromatography

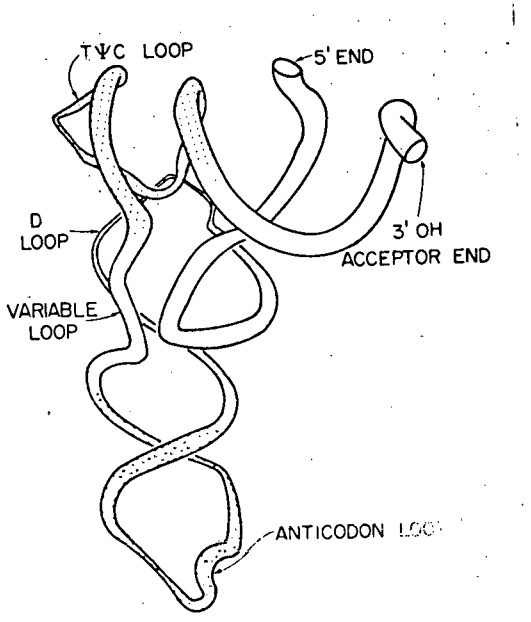


Figure 3 - Three-dimensional structure of yeast $tRNA^{Phe}$.

on BD-cellulose (30,31), DEAE-Sephadex (32), Sepharose 6B in acidic ammonium sulfate solution (33), "reversed phase" chromatography on RPC-5 columns (34) utilizing a combination of ionic and hydrophobic interactions to effect a separation of tRNAs, or the similar system employing the anion exchange polystyrene resin Aminex-A28 (35). Several affinity chromatography systems have been reported. Families of isoaccepting tRNAs can be enriched considerably by aminoacylation with the specific amino acid, derivatization with an aromatic acid function added to the α-NH_2 group of the aminoacyl ester, and chromatography on BD-cellulose (31). Extensively purified tRNA can be obtained by affinity chromatography on a column containing a bound, purified tRNA with an anticodon complementary to that of the desired tRNA species. For example, E.coli $tRNA^{Glu}$ (anticodon s^2UUC) is 19-fold enriched by passage through a column containing covalently bound yeast $tRNA^{Phe}$ (anticodon G_mAA) (36). Columns containing bound antibody raised against nucleoside Y or derivatives have been used in purifying $tRNA^{Phe}$ (37,38). Sepharose-bound lectins have been used in purifying tRNAs with a mannose-substituted Q base (39).

Also, the high resolving power of polyacrylamide gel electrophoresis has been utilized in one, two, or three dimensional systems to achieve purification of some tRNA species (40,41).

The strategy employed in our laboratory involves a multi-step enrichment by chromatography in BD-cellulose, Sepharose 6B in acidic ammonium sulfate reverse gradients, then repeated RPC-5 chromatography in several buffers until a single, symmetrical peak is obtained.

B.5. Sequence analysis of tRNA

The last several years have seen major advances in methods for sequence analysis of tRNAs that cannot be isolated in large quantities, or conveniently radiolabelled internally during cell growth. These "post-labelling" methods allow the complete sequence of a tRNA to be established using as little as 2-3 μg of purified material. An excellent discussion of many of these methods by Silberklang, Gillum, and RajBhandary is found in reference 42. Also, I have discussed tRNA sequencing methods in Chapter I of my M.Sc. thesis (43). Nevertheless, a comprehensive discussion of tRNA sequencing methods is relevant to the subject of this thesis.

The first complete tRNA sequence was completed by Holley and his group in 1965 (3). During this work, Holley worked out a general two-part strategy for sequence analysis of small RNAs. (i) The purified RNA is hydrolyzed to completion by nucleases with different specificities to generate distinct sets of fragments, which are then identified. Because of the different specificities of the enzymes used, some fragments can usually be aligned at this stage, since they will contain unique, overlapping sequences. (ii) The RNA is partially digested with the same enzymes to make larger fragments which are also separated and characterized. These are used to order the oligonucleotides obtained from complete digests. This general strategy, devised for sequence analysis of purified, non-radioactive RNAs, is discussed in a review by Holley (44).

A major limitation of methods used in Holley's work is the need for large amounts of purified RNA. Thus, sequence analysis could only be carried out on tRNAs that could be purified in bulk, for instance from yeast, wheat germ, or mammalian liver. The introduction by Sanger of methods for sequencing ^{32}P-labelled RNAs (45), adapting Holley's strategy, made

possible analysis of small amounts of tRNAs from cells that could be cultured conveniently in media containing ^{32}P-inorganic phosphate at high specific radioactivity.

The major advances in sequencing methodology in the past few years resulted from development of methods for enzymatic labelling of non-radioactive RNAs or oligonucleotides. A particularly useful enzyme in RNA sequence analysis is T4-induced polynucleotide kinase. Kinase-labelling of tRNA fragments generated by ribonuclease T_1, or ribonuclease A, separation, and identification of the [5' - ^{32}P]oligonucleotides is central to the strategy devised by RajBhandary and his colleagues for sequence analysis of small amounts of non-radioactive tRNA (42).

Conceptually similar methods for analyzing oligonucleotides labelled at the 3'-end with tritium have been developed by Randerath and co-workers (46). These methods, while relatively sensitive compared to the classical methods for analysis of non-radioactive RNAs, require much more purified RNA than methods employing phosphorus-32.

Sequence analysis of [5' - ^{32}P] or [3' - ^{32}P] RNAs or oligonucleotides has been reported. This can be accomplished by two-dimensional electrophoresis/homochromatography (47), or two-dimensional gel electrophoresis of partial enzymatic digests (48). Alternatively, methods have been developed for analysis of end-labelled RNAs by polyacrylamide gel electrophoresis of several partial, base-specific enzymatic hydrolysates (49,50). RNA labelled at the 3'-end may be subjected to partial, base-specific chemical modification, strand-scission at modified residues, and gel electrophoresis of the partial hydrolysates (51). These gel sequencing methods are especially useful in ordering fragments from limit ribonuclease digests.

The various methods mentioned above for sequence analysis of [^{32}P] tRNA or oligonucleotides have a common weakness. Modified nucleotides are frequently located in such a way that they cannot be identified directly as the [^{32}P] nucleotide. Rather, their identities must be inferred based on changes in electrophoretic behavior on addition or loss of the modified species in a nested set of homologous fragments, and on nuclease specificities. Thus, the method of Stanley and Vassilenko (52), and related procedures since developed in Randerath's (53) and Brownlee's (54) laboratories, has been an important advance.

The principle of the method is simple. If an intact, non-radioactive tRNA molecule is cleaved once in hot formamide, the resulting fragment containing the 5'-end of the molecule has a 5'-phosphate and 3'-cyclic phosphate while the other has a 5'-hydroxyl group and 3'-terminal CCA. The fragment with a 5'-hydroxyl group is a substrate for polynucleotide kinase, and can be radiolabelled with a [5'-^{32}P] phosphoryl group donated by [γ-^{32}P]ATP.

If a limited random hydrolysis of a purified tRNA is performed under conditions allowing zero or one cleavage per molecule, the set of [^{32}P] oligonucleotides obtained by polynucleotide kinase labelling is a nested set. These [5'-^{32}P] oligonucleotides, all containing the 3'-terminal CCA but varying at the 5'end, are separated according to size by polyacrylamide gel electrophoresis. Identification of 5'-terminal (^{32}P-labelled) nucleotides from the ordered set of oligonucleotides allows the nucleotide sequence, including modified nucleotides, to be "read" directly. This procedure will in principle allow identification, as [^{32}P]-nucleotides, of all nucleotides in the tRNA except those at the 5' and 3'-termini, and those

immediately 3' to a 2'-O-methylnucleotide. Furthermore, since very little material is required (as little as 0.5-1 μg), the method is applicable to tRNAs which can only be obtained in very small quantities. The Stanley and Vassilenko method is a powerful experimental approach to tRNA sequence analysis. The related methods of Gupta and Randerath (53) and Tanaka, Dyer, and Brownlee (54) are fast, accurate, and require little material. Their widespread use in sequence analysis of tRNAs in the future seems assured.

Combinations of the above post-labelling methods for sequence analysis have been applied to tRNAs from mitochondria (18,55), chloroplasts (56), neoplastic cell lines (41), and to minor cytoplasmic tRNAs present in small quantities such as suppressor tRNAs (57). Thus, systematic studies of sets of related tRNAs which may differ widely in amount are possible. Purity of material rather than quantity available is now the primary experimental obstacle in systematic studies of tRNA sequences.

C. Transfer RNA genes

Transfer RNA genes are an attractive model system for studies of gene expression. Transfer RNAs are involved in numerous cellular processes. Correct regulation of tRNA genes is therefore important to the cell. Because the gene product is small, stable, and readily characterized in very small amounts, the expression of these genes can be studied relatively easily at the molecular level. Mutant tRNA genes for informational suppressors have been identified in some organisms. Suppressor-carrying strains can be quite valuable in studies of cellular physiology.

Numerous studies on structure and organization of tRNA genes have been reported in the last few years. Some of these studies used genetic, electron microscopic, and RNA-DNA hybridization techniques. Quite recently, advances in this field have been aided tremendously by the burgeoning recombinant DNA technology.

Bacteriophage T_4 carries genes for eight tRNAs. These genes are located in two clusters within a single transcription unit. T_4 tRNA precursors have been extensively used to study tRNA biosynthesis (58,59).

In E. coli there are approximately 60 tRNA genes (58). Most of these genes are located in clusters at several locations on the chromosome, though genes for several tRNAs are found in the spacer region between 16S and 23S rRNA genes and are co-transcribed with the rRNAs (60,61). Probably the best-studied tRNA genes in E. coli are those for $tRNA^{Tyr}$. There are three genes coding for two nearly identical $tRNAs^{Tyr}$ differing only in the variable loop. Two identical genes for $tRNA_1^{Tyr}$ are located at a different site from the single gene for $tRNA_2^{Tyr}$ (62). Sequence determination of the $tRNA_1^{Tyr}$ precursor by Altman and Smith (63) led to the synthesis of a suppressor $tRNA_1^{Tyr}$ gene by Khorana and his group (64).

In eukaryotes, genes for a particular tRNA are generally redundant, existing in multiple copies. Organization and extent of duplication of tRNA genes varies considerably among eukaryotic species.

Yeast contain about 320-440 tRNA gene copies. Tyrosine and serine tRNA genes have been widely studied in the yeast S. cerevisiae since they can be converted to nonsense suppressors. There are eight genetically unlinked $tRNA^{Tyr}$ genes with identical tRNA coding sequences (65 and refs.

therein). These have been isolated in recombinant phage (66). The S. cerevisiae haploid genome contains at least eleven genes for $tRNA_2^{Ser}$; however, only a single gene for $tRNA_{UCG}^{Ser}$ is present (65,67). The organization of $tRNA^{Ser}$ genes is not really known. It is clear, however, that there is considerable variation in the number of genes for different isoacceptor tRNAs in yeast.

A different organization of tRNA genes is found in the clawed toad Xenopus laevis, which contains about 8000 gene copies. These are estimated by hybridization kinetics to consist of 43 tRNA sequences each reiterated about 200 times. There is extensive clustering of genes. Most of the tRNA genes are grouped together with spacer DNA as tandem repeat units extending up to 10^5 base pairs or more (68,69). Restriction mapping and DNA sequence analysis of a cloned 3.18 kb repeat unit reveals the presence of eight genes coding for tRNAs accepting phenylalanine, tyrosine, methionine (initiator tRNA; two copies), asparagine, alanine, leucine, and lysine (70,71). Only 70 base pairs separate the $tRNA^{Tyr}$ and $tRNA^{Phe}$ genes (71). There is no evidence, however, for co-transcription of more than one tRNA gene as is seen with bacteriophage T_4.

The human genome contains over 1000 tRNA genes, about 10-20 copies each of roughly 60 different genes (72). Consistent with this, Santos and Zasloff (73) find evidence for about 12 $tRNA_i^{Met}$ genes scattered on the human genome in Southern hybridization experiments. These workers sequenced two $tRNA_i^{Met}$ genes isolated in recombinant phage. Interestingly, one of the two genes contains the sequence GATCG, corresponding to the sequence GAUCG found in the TψC arm of all vertebrate $tRNA_i^{Met}$ sequences examined.

The other gene contains the sequence GATCT; however, no cytoplasmic tRNA of any sort has been found with a nucleotide other than G or A in this position. The significance of this observation is not known.

Mitochondrial and chloroplast chromosomes encode organellar sets of tRNAs distinct from those in the cell cytoplasm. The structures of mitochondrial tRNAs, based on RNA sequences or inferred from DNA sequences, are quite different from those of cytoplasmic tRNAs in several respects. The A+T richness of the mitochondrial genome extends to tRNA genes, in contrast to G+C rich cytoplasmic tRNA genes. Some mitochondrial tRNAs lack standard features of cytoplasmic tRNAs, such as the GG in the D loop; or have loops of unusually small size. A truncated mitochondrial serine tRNA and its gene lack the D arm altogether (74,75). Sequence analyses of mitochondrial DNA, tRNAs, and proteins demonstrate that different genetic codes are used in cytoplasm and mitochondria (76,77,18). The mammalian mitochondrial genome is very compact (78). Transfer RNA genes appear to be co-transcribed with rRNA and protein structural genes. They may function as signals for processing the single large polycistronic transcript from the circular genome into mRNA, rRNA, and tRNA (79).

A major result from the sequence analysis of eukaryotic DNA has been the finding that DNA and RNA sequences are not always co-linear. By the process of "splicing", internal RNA sequences can be removed from transcripts. In addition to protein structural genes (e.g. 80,81) and rRNA genes (82,83), tRNA genes from several eukaryotes contain intervening sequences located in the anticodon loop. Such split tRNA genes have been found in the genomic DNA of yeast (84-86,67), Drosophila melanogaster

(87), chicken (88), Xenopus laevis (71), and Zea mays (corn) chloroplasts (89). In most cases, the intervening sequences are small (about 8-40 nucleotides). The presumptive split tRNA genes identified by sequence analysis of Zea mays chloroplast DNA are particularly unusual. The intervening sequences in these tRNA genes are very large (over 800 base pairs). In addition, chloroplast tRNA genes are unusual because they are located in spacer sequences within a ribosomal RNA transcription unit. This organization is similar to that of rRNA operons of E. coli (61,90).

D. Transfer RNA and tRNA genes in Drosophila melanogaster

D.1. Transfer RNAs in Drosophila

D.1.a. Transfer RNA population

Isoacceptor tRNA populations from different developmental stages of Drosophila melanogaster were characterized by RPC-5 chromatography of aminoacyl-tRNA (91). The sensitive reversed phase chromatographic system allowed White et al. to identify 63 major and 36 minor isoaccepting tRNAs. Of these 99 tRNA peaks, approximately one-third undergo some quantitative change, relative to other isoacceptors, during development from first instar larvae to adults. Changes in peak sizes for major species $tRNA_5^{Thr}$, $tRNA_2^{Met}$ or $tRNA_3^{Met}$ were observed. Relatively few changes involve appearance or loss of a peak. Those that do concern a minor peak.

The elution profiles of asparaginyl, aspartyl, histidyl, and tyrosyl-tRNAs from different developmental stages were examined in detail (92). The relative proportions of certain chromatographically distinct

forms of these tRNAs change in similar fashion during the life cycle of wild-type flies. The δ and γ chromatographic forms were shown to differ only by the presence or absence, respectively, of a Q base derivative. The Q base presumably occupies the usual "wobble" position of the anticodon, since $tRNA^{His}_{1\gamma}$ contains G there (93). Wosnick and White (94) found that the relative amounts of δ and γ forms of $tRNA^{Tyr}$ change markedly dependent on growth conditions. The total amount of $tRNA^{Tyr}$ in adult flies is constant, however. Changes in the relative amounts of δ and γ forms are apparently a consequence of general metabolism rather than of genotype. This last point is relevant to reports by Jacobsen (95) and Twardzic et al. (96) that $tRNA^{Tyr}_{1\gamma}$ inhibits tryptophan pyrrolase, an inhibition relieved in flies carrying the suppressor of sable mutation. Their results have not proved to be reproducible, and apparent $tRNA^{Tyr}$ involvement in suppression of vermilion eye color mutations is probably artifactual (94,97).

The valine and serine tRNAs are the most extensively characterized isoacceptor families from Drosophila. The four valine codons GUN are read by three major species, $tRNA^{Val}_{3a}$ (GUA), $tRNA^{Val}_{3b}$ (GUG), and $tRNA^{Val}_{4}$ (GUA, GUC, GUU). The coding specificity of $tRNA^{Val}_{4}$ is a consequence of inosine, shown to be present by modified nucleotide analysis of the valine tRNAs (98). The extent of similarity of the three tRNAs was also examined by "fingerprint" analysis of complete ribonuclease digests (99). Quite recently, the nucleotide sequences of $tRNA^{Val}_{3b}$ (100) and $tRNA^{Val}_{4}$ (101) have been determined.

The six serine codons AGC, AGU, and UCN are read by major species $tRNA^{Ser}_{2}$ and $tRNA^{Ser}_{5}$ (AGC, AGU), $tRNA^{Ser}_{4}$ (UCG), $tRNA^{Ser}_{7}$ (UCU, and to a

lesser extent UCA and UCC), and $tRNA_6^{Ser}$ (specificity has not been determined) (102). Modified nucleotide analysis showed that $tRNA_2^{Ser}$ and $tRNA_5^{Ser}$ (AGC,AGU) are similar, and distinct from $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ (UCN). The nucleotide content of $tRNA_6^{Ser}$ is very similar to that of $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$, suggesting that it also reads UCN codon(s). Serine tRNAs responding to UCN codons in the ribosome binding assay contain the cytokinin isopentenyladenosine whereas those responding to AGC and AGU do not, consistent with serine tRNAs in other organisms (102). A detailed discussion of serine tRNAs is found in Chapter III below.

D.1.b. Nucleotide sequences of Drosophila tRNAs

The nucleotide sequences of a number of cytoplasmic tRNAs from Drosophila melanogaster have been determined. These include $tRNA_{3b}^{Val}$ (100), $tRNA_4^{Val}$ (101), $tRNA_4^{Glu}$ (103), $tRNA_{1\gamma}^{His}$ (93), $tRNA_2^{Lys}$ (104), $tRNA_i^{Met}$ (105), and $tRNA^{Phe}$ (106). These tRNAs are generally quite similar to the equivalent vertebrate tRNAs, indicating strong conservation of sequences over evolutionary time periods. They are unremarkable otherwise, fitting very well with the generalized tRNA structure of Rich and RajBhandary (8).

D.2. Organization and structure of tRNA genes in Drosophila

D.2.a. Hybridization of tRNAs and genes to polytene chromosomes in situ

Hybridization to polytene chromosomes in situ is a powerful method for localizing the genes for purified tRNAs in Drosophila. Stained polytene chromosomes from salivary glands show a reproducible banding pattern that is readily visible under the light microscope. Each of the roughly 500

genome copies in the polytene chromosomes contains about 600 genes for the 60 tRNA species (107). Purified ^{3}H or ^{125}I-labelled tRNA can be hybridized to these chromosomes, and tRNA genes located by the appearance of silver grains over specific chromosomal bands after autoradiographic exposure. The first gene localization of this sort was reported by Grigliatti et al. (108). However, despite extensive precautions, even the highly purified $tRNA_5^{Lys}$ used in those experiments proved to be sufficiently contaminated that the localization was incorrect (109). Through improvements in purification and hybridization conditions, the method has been refined to an extent that has allowed localization of genes for over twenty tRNAs (109,110,111, 112). This number is over one-third of the total major tRNA species in Drosophila (91).

Two major conclusions were reached based on in situ hybridization of purified tRNAs. First, tRNA genes are spread about the chromosomes. Sites of hybridization that presumably contain tRNA genes have been found on every chromosomal arm except the very small chromosome 4 (109). Second, tRNA genes seem to be arranged as dispersed clusters. Individual tRNAs generally hybridize to more than one site on polytene chromosomes. These sites may contain more than one gene. For example, $tRNA_{3b}^{Val}$ hybridizes to regions 84D, 92B, and 90BC on the right arm of chromosome 3 (3R). Based on the numbers of grains over these sites, it was concluded that they contained 5, 4, and 1 genes (or multiples thereof), respectively (113). In situ hybridization can be performed using radiolabelled recombinant plasmids containing Drosophila DNA. Three plasmids selected by hybridization with ^{125}I-$tRNA_{3b}^{Val}$ each hybridized to single chromosomal sites. These

were 84D and 90BC, the same as sites hybridizing purified tRNA. Two plasmids hybridized to the minor site at 90BC, showing that grains at this site in tRNA experiments were not artifactual (114). Similar results were obtained on hybridization of $tRNA_4^{Val}$, $tRNA_4^{Ser}$, $tRNA_7^{Ser}$, and recombinant plasmids selected with those tRNAs, to polytene chromosomes in situ (109, 114).

D.2.b. Analysis of recombinant plasmids carrying Drosophila tRNA genes

D.2.b.i. Genes in the 42A region: the structure of a tRNA gene cluster

A major cluster of tRNA genes identified by hybridization of tRNAs to polytene chromosomes in situ is in the 42A region on chromosome 2R (109,112). The tremendous advances in recombinant DNA technology over the past several years have made possible the isolation and characterization, even to the nucleotide level, of large segments of DNA from this chromosomal region. The first recombinant plasmid isolated containing Drosophila tRNA genes, pCIT 12, derived from the 42A region (115). DNA sequence analysis of all tRNA-coding segments of this plasmid was carried out by Soll and his co-workers (116). This group found eight tRNA genes in the 9.3 kilobase pair (kb) insert of pCIT 12: three for $tRNA^{Asn}$, three for $tRNA_2^{Lys}$, one for $tRNA_2^{Arg}$, and one for $tRNA^{Ile}$. This analysis was extended by Yen and Davidson (117), who isolated overlapping recombinant phage spanning 94 kb of sequence derived from the 42A region. By restriction enzyme mapping techniques and some DNA sequence analysis, they showed that a region of 46 kb (including pCIT 12 sequences) contained 18 tRNA genes: eight for $tRNA^{Asn}$, four for $tRNA_2^{Arg}$, five for $tRNA_2^{Lys}$, and a single $tRNA^{Ile}$ gene. These genes are irregularly spaced and are not all transcribed from

the same DNA strand. In all cases where DNA sequence was determined, redundant genes at 42A hybridizing a particular tRNA have identical sequences.

D.2.b.ii. Other tRNA gene clusters

Transfer RNA genes from gene clusters located at other chromosomal sites than 42A have also been analyzed after isolation in recombinant vectors. These include genes for $tRNA_2^{Lys}$ (118), $tRNA^{Glu}$ (119), $tRNA^{Leu}$ and $tRNA^{Ile}$ (87), $tRNA^{Gly}$ (120), $tRNA_i^{Met}$ (121), and $tRNA_4^{Val}$ (122). The organization of these clusters is somewhat like that seen at region 42A. Multiple tRNA gene copies are present, some close together. Two clusters containing five $tRNA^{Glu}$ genes are present within 4 kb of sequence deriving from chromosomal site 62A (119). Robinson and Davidson found seven genes for two tRNAs in a 2.5 kb sequence from region 50AB (87). Four $tRNA_2^{Lys}$ genes are present in a 3 kb segment of DNA from region 42E (118). Thus, $tRNA_2^{Lys}$ genes from regions 42A and 42E, the two major sites of $tRNA_2^{Lys}$ hybridization in situ (109), have been isolated.

On the other hand, tRNA genes in a cluster may be more widely spread. Three $tRNA_i^{Met}$ genes from region 61D are separated on the genome by at least 15 kb (121). Thus, the organization of tRNA gene clusters is apparently quite flexible.

D.2.c. Are redundant tRNA genes identical?

As described above, the DNA sequences of a number of tRNA genes from D. melanogaster have been determined. In most cases where the DNA sequences are known, the redundant tRNA genes are identical. Most of

the redundant genes for $tRNA^{Asn}$, $tRNA_2^{Lys}$, and $tRNA_2^{Arg}$ from region 42A have been sequenced and are identical (116,117). A $tRNA_2^{Lys}$ gene isolated from 42E in the genome is identical to genes from the 42A region (123,118). Five genes for $tRNA^{Ile}$ from region 50AB are all identical to the single $tRNA^{Ile}$ gene from 42A in pCIT 12 (87,116). Two $tRNA^{Leu}$ genes from 50AB are the only "split" tRNA genes identified from D. melanogaster to date (87). These genes encode the same mature $tRNA^{Leu}$, though the intervening sequences differ.

However, it is becoming clear that not all genes hybridizing a particular tRNA have identical sequences. Five $tRNA^{Glu}$ genes were found on plasmids derived from the 62A region. Four of the genes are identical; the fifth differs by a single C to T transition at the fourth nucleotide of the mature tRNA sequence (119). This represents one of two instances thus far where non-identical genes for a tRNA originating at the same chromosomal site have been reported. Genes for $tRNA_i^{Met}$ with sequences matching and differing from the known tRNA sequence are present at region 61D (121). Non-identical sequences isolated by hybridization with $tRNA_4^{Val}$ (122) and $tRNA_5^{Lys}$ (124) have been found recently. These differing genes derive from distinct chromosomal sites. Thus, while most sequences hybridizing with a single tRNA species may be identical, not all are. The possible significance of such non-identity of tRNA genes is discussed in detail in Chapter IV.

D.3. Expression of Drosophila tRNA genes

D.3.a. Expression of $tRNA^{Val}$ genes in mutant flies

Genes for $tRNA_{3b}^{Val}$ are found in two major clusters on D. melanogaster polytene chromosomes (109). Dunn et al. examined expression of genes for

$tRNA_{3b}^{Val}$ in mutants carrying duplications or deficiencies of DNA from one of the two sites (113). Flies with a duplication of the 84D region had increased amounts of $tRNA_{3b}^{Val}$ and of $tRNA^{Val}$. The increase in $tRNA_{3b}^{Val}$, about 30%, was proportional to the increase in gene number, based on a value of nine genes total, five at 84D, in the wild type haploid genome. A corresponding proportional decrease in the amount of $tRNA_{3b}^{Val}$ was seen in flies deficient for the 84D region. However, the total amount of $tRNA^{Val}$ present was the same as for wild type flies. This suggests that a homeostatic mechanism for maintaining the total amount of $tRNA^{Val}$ exists.

D.3.b. Transcription of Drosophila tRNA genes

The existence of internal, split promoters for eukaryotic, RNA polymerase III-transcribed 5S rRNA and tRNA genes has been reported recently (125,126,70). This conclusion is based on studies of transcription of systematically altered recombinant DNA containing genes from Xenopus laevis, in a homologous in vitro transcription system derived from extracts of Xenopus germinal vesicles. Specific transcription can also be achieved "in vivo" by microinjecting recombinant plasmids into nuclei of intact Xenopus oocytes.

The Xenopus in vivo and in vitro systems have been used to study transcription of tRNA genes from heterologous sources isolated in recombinant plasmids. The primary transcript of a Drosophila $tRNA_2^{Arg}$ gene, including the 5'-terminal nucleotide, was thus obtained and characterized (127). Experiments using Xenopus germinal vesicle extracts indicated that the mature tRNA coding sequence is sufficient for accurate transcription initiation on

Drosophila $tRNA_2^{Lys}$ genes (123). In the same experiments, DeFranco, Schmidt, and Soll showed that 5' gene-flanking sequences modulated the level of specific gene transcription.

It appears that conclusions based on transcription of tRNA genes from heterologous sources in *Xenopus* systems are valid and generally applicable. Yeast $tRNA^{Tyr}$ genes containing intervening sequences are transcribed, the transcripts processed and partially modified quite accurately in *Xenopus* oocytes (128). Transcription of *D. melanogaster* and *B. mori* tRNA genes in homologous sytems is the same as in *Xenopus* germinal vesicle extracts (129,130).

Two major conclusions are suggested by studies on transcription of tRNA genes from eukaryotes, including *Drosophila*, up to the present time. First, accurate initiation of tRNA gene transcription depends on an internal, split promoter (123,70). Second, the level of transcription is modulated by flanking sequences 5' to the gene (123,130).

A primary purpose for the recent intensive DNA sequence analysis of cloned tRNA genes is to have well-characterized templates for studies of gene expression. The DNA sequences of many tRNA genes are now known. Thus, we can expect to see many studies relating to the control of tRNA gene transcription in the next several years.

Chapter II

MATERIALS AND METHODS

A. RNA Sequence Analysis

Materials: Chemicals used were reagent grade commercial stocks. Urea (ultra-pure) for partial enzymatic hydrolyses of RNA and E.coli tRNA were from Schwarz-Mann. Nucleoside 5'-phosphates for use as chromatographic standards and ATP were from P-L Biochemicals. Cytidine (2' + 3')-phosphate was from Calbiochem. Thin layer chromatography plates (Macherey-Nagel CEL 300 DEAE, Macherey-Nagel CEL 300 PEI, and cellulose) were purchased from Brinkmann. Cellulose acetate membrane strips (3cm x 55 cm) were from Schleicher and Schuell. Carrier-free $H_3{}^{32}PO_4$ was obtained from New England Nuclear or from Amersham. Sodium 3-phosphoglycerate (Grade I) was from Sigma Chemical Co. [γ-^{32}P] ATP was either purchased from Amersham Radiochemicals (specific activity at least 2000 Ci/mmole) or prepared by the method of Glynn and Chappell (131) as described by Maxam and Gilbert (132) to a specific activity of 1000-1500 Ci/mmole. Unless stated otherwise, most procedures were carried out in 1.5 ml or 1.9 ml capped, conical polypropylene tubes (Bio-Rad or Evergreen).

Enzymes were obtained from the following sources: Boehringer-Mannheim (calf intestinal phosphatase), Calbiochem (ribonucleases T_1, T_2, U_2, A, nuclease P_1 of Penicillium citrinum, and hexokinase), Enzo Biochemicals (RNase Phy I from Physarum polycephalum), New England BioLabs (T_4

polynucleotide kinase and RNA ligase (RNase-free)), Sigma Chemical Co. (3-phosphoglycerate phosphokinase), and Worthington Biochemicals (rabbit muscle glyceraldehyde 3-phosphate dehydrogenase).

Individual tRNA species were purified by Dr. Ian Gillam by established procedures and provided for RNA sequence analysis.

Methods

A.1. Nucleotide content analysis

This was performed essentially as described (42), with the following modifications. After kinase-labelling and conversion of excess [γ-^{32}P] ATP to glucose-6-$^{32}PO_4$, the nuclease P_1 reaction was carried out in the same Tris-buffered solution, since the 3'-phosphatase activity of nuclease P_1 is near its maximum in the pH range 7.5-8.0 (133). To quantitate the nucleotides, spots on the two dimensional cellulose plates located by autoradiography were scraped from the backing, collected by suction in pipettor tips plugged with glass wool, eluted with H_2O, and Cerenkov cpm determined.

A.2. Synthesis of [5'-^{32}P] cytidine 5', 3'-bisphosphate (*pCp)

The reaction mix (10 µl) contained 1.2 mM Cp (2' + 3'), 30 mM TrisHCl (pH 8.3), 10 mM $MgCl_2$, 10 mM dithiothreitol, 30-40 µM [γ-^{32}P]ATP (specific activity $\geq$ 300 Ci/mmole), and 1-2 units of T_4 polynucleotide kinase (100-200 U/ml). The reaction (40-60 min. at 37°) was stopped by heating 1 min. at 100°. Storage was at -20°. [γ-^{32}P] ATP was assumed to be quantitatively converted to *pCp (starting Cp present in 30-40 fold excess over

ATP); thus, the specific activity of the *pCp is the same as that of input [γ-^{32}P] ATP. Excess Cp(2' + 3') does not seem to inhibit the subsequent RNA ligase-catalyzed labelling reaction. Therefore *pCp was prepared and used in the labelling reaction without further purification.

A.3. Synthesis of [3'-^{32}P] tRNA by RNA ligase-catalyzed addition of *pCp

This reaction was performed as described (134). The reaction mix (30 µl) contained 1-2 µg of purified tRNA, 50 mM HEPES (pH 8.3), 10 mM $MgCl_2$, 3.3 mM dithiothreitol, 5 µM ATP, 1 µl *pCp mix (from 10 µl total; ≃ 1 µM *pCp final concentration), 10% DMSO, 15% glycerol, and 1 µl of RNA ligase. Labelling reactions were for 16-24 hours at 4°. These were stopped by adding an equal volume of 90% formamide containing 0.067% (w/v) xylene cyanol FF (XC) and bromophenol blue (BB). [3'-^{32}P] RNA with a *pCp adduct was purified by electrophoresis on thin, denaturing polyacrylamide gels (135), located by autoradiography, and eluted (52). Based on radioactivity recovered from full-sized tRNA products after excision of the appropriate gel slice, input tRNA was radiolabelled and recovered with 20-30% efficiency.

A.4. De-phosphorylation of tRNA

De-phosphorylation was carried out essentially as described (42). The reaction mix (10 µl) contained 2-3 µg of tRNA, 20 mM TrisHCl (pH 8.0), 0.2 mM EDTA, and 0.024 units of calf intestinal phosphatase. Reactions were for 40-75 min. at 55°, followed by phenol extraction and ether extraction as described (42). Samples were then dried in a dessicator containing P_2O_5 under aspirator vacuum.

A.5. Polynucleotide kinase-catalyzed synthesis of [5'-^{32}P]tRNA or oligonucleotides

This reaction was carried out in the same tube as de-phosphorylation. The procedure was based on that of Lillehaug and Kleppe (136). The reaction mix (10 µl) contained (5'-OH)tRNA or oligonucleotides; 60 mM TrisHCl (pH 8.0), 9 mM $MgCl_2$, 10 mM dithiothreitol, 40-50 µM [γ-^{32}P]ATP (specific activity ≥ 400/mmole), and 1-2 units polynucleotide kinase (100-200 U/ml). The reactions were for 30-40 min. at 37°, and were stopped by adding an equal volume of formamide/dye mix (13:1:1, formamide:1% XC:1% BB), heating 1 min. at 100°, and chilling on ice. Samples were stored at -20% until use.

A.6 Purification of [5'-^{32}P] RNA

[5'-^{32}P]RNAs were purified by polyacrylamide gel electrophoresis (135), located by autoradiography, and eluted as described (52). The labelling reaction is inefficient, yielding about 250,000 Cerenkov cpm/µg input tRNA in the full-sized tRNA (< 10% yield). The yield for oligonucleotides may be considerably higher than for intact tRNA, but varies with individual substrates and with the particular lot of enzyme used.

A.7. "Gel read-off" sequence analysis of [5'-^{32}P] or [3'-^{32}P]RNAs

Partial base-specific hydrolyses of [^{32}P]RNA plus carrier tRNA with RNase T_1 and RNase A were performed essentially as described by Donis-Keller et al. (49) except that reaction volumes were 10 µl instead of 20 µl,

marker dyes were added after the reactions were stopped, and serial dilutions of enzymes were not normally performed. Random hydrolysis in hot formamide, or base-specific hydrolyses with RNase U_2 and RNase Phy I were performed as described (50). Specific details on ratios of enzyme to substrate are found in figure legends. Electrophoresis of the hydrolysates was in adjacent slots of thin, denaturing polyacrylamide gels (40 cm x 20 cm x 0.05 cm, slots 1 cm x 1 cm x 0.05 cm). The concentrations of acrylamide used in these gels were 12%, 20%, and 25% (w/v). All gels contained 20 gm acrylamide per gm of methylene-bis-acrylamide, 7M urea, 90 mM Tris-borate (pH 8.3), and 2.5 mM EDTA. For polymerization, ammonium persulfate was added to 0.05% (w/v) and TEMED to 0.1% (v/v). Polymerization was normally completed within 15 min. Electrophoresis was carried out at constant power (about 20 watts), monitoring the run by the migrations of the marker dyes XC and BB. Bromophenol blue comigrates with oligonucleotides of length 4-5 (25% gel), 6 (20% gel), or 13-15 (12% gel), [^{32}P] oligonucleotides were then located by autoradiography, and the order of nucleotides "read off" from the autoradiographs.

A.8 Sequence analysis of tRNA by the method of Stanley and Vassilenko (52)

A.8.a. Purification of tRNA by polyacrylamide gel electrophoresis

To obtain clear results sequencing by the Stanley and Vassilenko method, the tRNA sample should be intact and free of significant amounts of any single contaminating RNA species. In analysis of $tRNA^{Ser}_{2b}$, further purification of the starting material was required. This was accomplished

by polyacrylamide gel electrophoresis of 10 μg of $tRNA_{2b}^{Ser}$ for 40 hr. at 300V on a 20% gel containing 4M urea, 90 mM Tris-borate (pH 8.3), and 2.5 mM EDTA (137). Gel dimensions were 40 cm x 20 cm x 0.15 cm, with slots 1 cm x 1 cm x 0.15 cm. The full-sized tRNA was located under UV light after staining the gel for 10 min. with buffer containing 0.1 μg/ml ethidium bromide. (It is worth noting here that while some protocols call for staining the gel with methylene blue (137) or toluidine blue (54), these dyes must be used with special care as both catalyze cleavage of DNA and RNA at guanosine residues in the presence of visible light and molecular oxygen (138,139). This can result in very high backgrounds in subsequent sequence analysis.) RNA was eluted from the excised gel slice in 0.40 ml of elution buffer (0.5 M NH_4OAc, 10 mM $Mg(OAc)_2$, 1 mM EDTA, 0.1% SDS) overnight at 4°. The tRNA eluted was recovered by ethanol precipitation (2.5 volumes of 95% ethanol, 1 hr. on solid carbon dioxide) and centrifugation for 4 min. at room temperature in an Eppendorf centrifuge. The supernatant liquid was discarded. The pellet was washed with 95% ethanol (-20%) and recovered as above, then dried briefly under aspirator vacuum.

A.8.b. Partial hydrolysis of the tRNA

A variety of hydrolysis conditions were used, in formamide solutions and in several buffered aqueous solutions. Good results were obtained when hydrolyses were performed in 70-100% formamide. Of several aqueous systems tried, hydrolysis in 10 mM NH_4OAc (pH 4.6) worked best, giving relatively uniform intensity of labelling and low background.

Hydrolyses in formamide were carried out in the following way. Samples of tRNA (0.5-6 μg, dry or in a small volume of water) in a conical polypropylene tube were mixed with formamide to give 70-100% formamide in 10-20 μl total. Hydrolysis was for 10-15 min. in the sealed tube in a boiling water bath, and was stopped by chilling the tube on ice. One volume of 0.6M NaOAc, then five volumes of 95% ethanol were added, mixing by inversion after the latter. Samples were stored at -70% for at least two hr., or overnight at -20°, then the RNA was collected by centrifugation in an Eppendorf centrifuge (4 min. at room temperature). The supernatant liquid was carefully removed, discarded, and the pellet dried under aspirator vacuum.

Hydrolyses of 0.5-2 μg of tRNA in 8-10 mM NH_4OAc (pH 4.6) were for 1.5 - 3 min. in a boiling water bath in a sealed capillary tube. The capillary was then quickly chilled in an ice-water bath, its contents expelled into a conical polypropylene tube and dried under aspirator vacuum.

A.8.c. Labelling hydrolysis products with ^{32}P by polynucleotide kinase

The labelling reaction was carried out in the tube contained the dried hydrolysate. This was done as described above for synthesis of [5'-^{32}P] tRNA or oligonucleotides, unless the [γ-^{32}P]ATP was a commercial stock of high specific activity ($\geq$ 2000 Ci/mmole). In that case, the concentration of ATP was 5-10 μM.

A.8.d. Separation of [5'-^{32}P]tRNA fragments by polyacrylamide gel electro-electrophoresis

[5'-^{32}P] oligonucleotides generated from tRNA by partial hydrolysis and post-labelling were separated according to size by electrophoresis on denaturing polyacrylamide gels, as described above for the gel "read-off" sequencing method. From a 20 µl sample (50% formamide, containing SC and BB), aliquots of 7-10 µl were loaded into 1 cm x 1 cm x 0.05 cm slots and run for varying lengths of time on 20% or 12% gels in order to cover the entire tRNA sequence. (For example, see Fig. 1 of ref. 53.) Anion exchange paper was placed beneath the gel in the lower buffer chamber to trap $^{32}PO_4$ and unreacted [γ-^{32}P]ATP (42). The [^{32}P]RNA bands were located by autoradiography, the bands excised from the gel in order of increasing oligonucleotide length, and each oligonucleotide eluted overnight at 4° in 0.40 ml 0.3M NaCl, 0.1% SDS (52).

Analysis of 5'-terminal (^{32}P-labelled) nucleotides from oligonucleotides

A.8.e. Analysis of 5'-termini as *pNp's

Analysis of 5'-terminal *pNp's was carried out by one of three related protocols in sequence analysis of (1) $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$, (2) $tRNA_5^{Lys}$, and (3) $tRNA_{2b}^{Ser}$. (1) Hydrolyses of isolated [5'-^{32}P] oligonucleotides were with 0.2M NaOH. Such hydrolyses and subsequent identification of the *pNp's by chromatography on PEI-cellulose were carried out precisely as described in Chapter 2 of reference 43.

There are certain shortcomings to alkaline hydrolysis of the [^{32}P] oligonucleotides. The products of alkaline hydrolysis are nucleoside

5', 3'-bisphosphates and nucleoside 5', 2'-bisphosphates; the latter are resistant to the 3'-phosphatase activity of nuclease P_1 and cannot be converted to nucleoside 5'-phosphates (pN's) for further characterization (133). Also, some modified nucleosides are labile in alkaline solutions. For example, 4-acetylcytidine loses its acetyl substituent with a half-time of 6.8 min. at 37° in 0.1M NaOH (4); the purine ring of 7-methylguanosine is cleaved at high pH (140); and 1-methyladenosine is converted by ring-opening and rearrangement to 6-methyladenosine under such conditions (42). Due to such considerations, mild enzymatic hydrolysis conditions were used in later sequence analysis.

(2) Part or all of the ^{32}P-labelled material recovered from each gel slice (as above) and 30µg of carrier E.coli tRNA was precipitated with ethanol and collected by centrifugation in an Eppendorf centrifuge (4 min. at room temperature). After discarding the supernatant liquid, the RNA-containing pellet was air-dried, then digested overnight at 37° with 0.05 unit of RNase T_2 in 10 µl 20 mM NH_4OAc (pH 4.6). An aliquot of each ribonuclease digest (2 µl) was spotted at 1 cm intervals at the origin on a 20 cm x 20 cm PEI-cellulose plate (prepared in the formate form as described in ref. 141). The origin was washed with H_2O, and the plate developed with 0.8M $(NH_4)_2 SO_4$, 2 mM EDTA at room temperature. The *pNp's (3'-phosphorylated, but not 2'-) were located by autoradiography of the developed thin layer plate.

(3) [^{32}P] RNAs recovered by elution from gel slices were recovered by ethanol precipitation and centrifugation as above, together with 10 µg

E. coli tRNA carrier. After discarding the supernatant liquid, the pellet was air-dried, then dissolved in 10 µl of 10 mM NH_4 OAc (pH 4.6) containing 0.2 unit RNase T_1, 0.2 unit RNase A, and 0.1 unit RNase T_2. The RNA was digested overnight at 37°. The hydrolysates were air-dried, dissolved in 5 µl H_2O, and 1 µl of each sample applied at 1 cm intervals on 20 cm x 20 cm PEI-cellulose thin layer plates (formate form). The origins of the plates were washed with H_2O, the plates developed in 0.8 M $(NH_4)_2$ SO_4, 2 mM EDTA at room temperature, and the nucleotides located by autoradiography.

A.8.f. Analysis of 5'-termini as *pN's

5'-terminal (^{32}P-labelled) nucleotides thought likely to be modified, by virtue of chromatographic behavior as the *pNp on PEI-cellulose (above) or position in the nucleotide sequence, were analyzed further as the *pN's. The *pN's were generated by nuclease P_1 digestion of either intact [5'-^{32}P] oligonucleotides (endonuclease activity) or of *pNp's (3') resulting from RNase T_2 digestion of the oligonucleotides (3'-phosphatase activity of nuclease P_1) (133). In the former case, the [^{32}P]oligonucleotides were collected together with carrier tRNA by ethanol precipitation and centrifugation (as described above), and digested in 10 µl 10 mM NH_4OAc (pH 4.6) containing 0.1 µg nuclease P_1 for at least 2 hr. at 37°. In the latter case, 0.1 µg nuclease P_1 was added directly to the ribonuclease T_2 digestion mix. The conversion reaction (*pNp to *pN) was for at least 2 hr. at 37°.

Different hydrolysis conditions were used when the 5'-terminal nucleotide was deemed likely to be ribose-methylated (*pN_m). In this case, 2-6 µg of nuclease P_1 was added (10 µl of 10 mM NH_4OAc, pH 4.6), and digestion was for at least 7 hr. at 37° (43, 142).

A.8.g. Characterization of *pN's by thin layer chromatography

The chromatographic mobilities of a variety of modified nucleoside 5'phosphates have been reported for several different solvent systems (42, 143). In the work presented in this thesis, chromatographic mibilities on cellulose thin layer plates were determined for *pN's in solvent A (66 ml isobutyric acid: 1 ml conc. NH_4OH: 32 ml H_2O: 1 ml 0.2 M EDTA, pH 8.5) or solvent B (100 ml 0.1 M Na_2HPO_4, pH 6.8: 68 gm $(NH_4)_2$ SO_4: 2 ml 1-propanol: 1 ml 0.2 M EDTA, pH 8.5) (42). In many cases, the *pN's were chromatographed in two dimensions, developing in the first dimension with solvent A and in the second with solvent B, as described (42,43). Nucleotides were identified on the basis of chromatographic mobilities in one or two dimensions relative to nucleoside 5'-phosphate standards run on the same plates. Comigration of *pN with a nucleotide standard in both solvent systems was generally considered sufficient to identify the *pN. In other cases chromatographic mobilities in solvents A and B, taken together with the mobility of the *pNp on PEI-cellulose in 0.8 M $(NH_4)_2$ SO_4 and ribonuclease specificities in gel read-off experiments, allowed identification of the modified nucleotide.

A.8.h. Characterization of sulfur-containing nucleotides

Nucleotides thought likely to be 2-thio derivatives of the standard bases were tested for sensitivity to cyanogen bromide (CNBr) as described (144,145). In mild alkaline solutions, CNBr will form cyanido derivatives of nucleotides with exocyclic thio-modifications, such as s^2U and s^4U. Such cyanido nucleotide derivatives are unstable in acidic solutions (144).

The *pN sample to be tested (at least 2000 Cerenkov cpm) was dissolved in 20 µl of 0.1 M NH_4HCO_3, pH 9.0. All subsequent steps were carried out in a fume hood. As the control, a 5 µl aliquot was removed prior to treatment with CNBr. The modification reaction was started by adding 2 µl of CNBr (25 mg/ml in ethanol) to the remaining 15 µl of *pN sample. The reaction was for 10 min. at room temperature. After 10 min., 10 µl was removed from the reaction mix into 100 µl of 0.1 M HCl, while the remainder was frozen on solid CO_2. Aliquots of 55 µl were taken from the HCl solution and frozen on solid CO_2 after 4 hr. and 22 hr. at room temperature. All samples were dried in a dessicator under aspirator vacuum, in a fume hood. The dried samples were dissolved in 2 µl H_2O, then applied to cellulose thin layer plates. The chromatograms were developed in either solvent A or solvent B, as described above. The [^{32}P]-nucleotides were located by autoradiography. Nucleotide standards were located under UV light.

A.9. Two-dimensional RNA sequencing: a variation on Stanley/Vassilenko

Two-dimensional RNA sequence analysis was carried out as described by Tanaka, Dyer, and Brownlee (54) except that electrophoresis was in polyacrylamide gels 40 cm in length rather than 80 cm, and transfers of separated oligonucleotides were to 20 cm x 20 cm DEAE-cellulose plates rather than to 20 cm x 40 cm plates.

Preparation of [^{32}P]RNA samples by limited hydrolysis in hot formamide, post-labelling, and subsequent polyacrylamide gel electrophoresis were as described above (Sequence analysis of tRNA by the method of Stanley and Vassilenko). DEAE-cellulose plates were treated before use, washing by ascending chromatography in 40 mM EDTA, pH 4.7, then again in H_2O in the

same direction. (Note that the pH of the EDTA solution used here is important.) A segment of the polyacrylamide gel 20 cm x 1.5 cm was excised using a scalpel and removed by adhesion to "used" X-ray film, then transferred to the wetted (with 1M NH_4OAc, pH 4.6) origin of a DEAE-cellulose plate, as described (54). A glass plate was placed over the transfer, the transfer and glass plates wrapped tightly in cellophane, and weights (several pounds of lead blocks) placed over the origin where the gel strip lay. Transfers were left for 16-24 hr. at room temperature (transfer efficiencies of about 50% from 20% gels or about 75% from 12% gels were observed). They were then washed in H_2O and dried. The [5'-^{32}P] oligonucleotides were then digested by ribonuclease <u>in situ</u>. About 100 µl of RNase mix (0.10 M NH_4OAc, pH 4.6, containing 0.2 U RNase T_1/µl, 0.2 U RNase A/µl, and 0.1 U RNase T_2/µl) was applied along the origin on the DEAE-cellulose plate with a pipettor (taking care not to touch the surface of the plate, which is easily damaged, with the pipettor tip). The plate was wrapped in Saranwrap, clamped between glass plates and incubated at 37° for 2-4 hr. The origin of the DEAE-plate was washed for about 10 min. at room temperature in methanol, and the plate air-dried. Electrophoresis in the second dimension was in the same direction as the pre-washing in EDTA. Electrophoresis was performed using an acid buffer composed of 8% acetic acid, 2% formic acid, 5 mM EDTA (pH 2.3). The DEAE-cellulose plate was wetted by blotting with Whatman 3 MM filter strips soaked in the acid buffer, covering the origin last. Wicks from the buffer chambers were also 3 MM filter paper, thoroughly wetted in the running buffer. Better results were obtained when the plates were left exposed to the air rather than covered with Saranwrap. Electrophoresis of *pNp's generated by <u>in situ</u> RNase hydrolysis was for 2.5-3 hr. at 300 V on a water-cooled Savant flatbed electrophoresis apparatus. At this time, xylene

cyanol marker dye spotted at the origin had migrated about 1.2-1.5 cm. Individual *pNp's were located by autoradiography.

A.10. "Wandering spot" analysis of [5'-^{32}P] or [3'-^{32}P]tRNA: two-dimensional electrophoresis/homochromatography

The methods used were essentially those described by Jay et al. (47), except that partial hydrolyses of end-labelled oligonucleotides were in hot formamide (rather than the enzymatic methods used in refs. 42, 47, and 48). A sample of ^{32}P-end labelled RNA was partially hydrolyzed in 70-100% formamide for 60 min. (5'-labelled RNA) or 20 min. (3' *pCp-labelled RNA) at 100°. Electrophoresis on a water-cooled Shandon flatbed electrophoresis apparatus was performed in the following way. A strip of cellulose nitrate was wetted on both sides by drawing it twice through a trough containing 5% acetic acid, 7M urea (pH 3.5). Excess buffer was removed from the strip by running a finger over it lengthwise, then blotted with tissue paper. A section of the membrane strip about 10 cm from the cathode buffer chamber was thoroughly blotted with tissue paper before applying 1-4 µl of hydrolysate containing a minimum of 5000 Cerenkov cpm. A small volume of marker dye solution (0.33% (w/v) of xylene cyanol FF, orange G, and acid fuchsin) was spotted on either side of the [^{32}P]hydrolysate at the origin. After sample and dyes had soaked into the cellulose nitrate strip, both ends of the strip were covered (above and below) by Whatman 3MM filter paper wicks from the buffer chambers, thoroughly wetted with running buffer (5% acetic acid, adjusted with pyridine to pH 3.5). Membrane strips and wicks were covered with mylar film, then with glass plates. Electrophoresis was for 30-45 min. at 3000V with water-cooling, until the orange and blue dyes were

separated by 12-15 cm. Oligonucleotides were then transferred from the membrane strips to 20 cm x 40 cm DEAE-cellulose plates as described (146). The plates were washed with H_2O to remove urea along the transfer line, then air-dried. Homochromatography was carried out essentially as described by Jay et al. (47).

The origins on the plates were wetted by blotting with 3MM filter paper strips in H_2O, then the plates were developed with homomixture 5 (47) at 65° until the solvent front reached the top (10-12 hr.). The ^{32}P-labelled oligonucleotides were located by autoradiography.

B. DNA Sequence Analysis

Materials

Chemicals used were generally reagent grade commercial stocks. Hydrazine was from Eastman-Kodak. Dimethylsulfate was obtained from Aldrich ("Gold Label" reagent). Piperidine, from Fisher Chemical Co., was re-distilled before use. [α-^{32}P] deoxynucleoside 5'triphosphates (specific radioactivity at least 2000 Ci/mmole) were obtained from Amersham Radiochemicals. E. coli tRNA was from Schwartz-Mann. Agarose (electrophoretic purity) was obtained from Bio-Rad. Most of the restriction enzymes used were obtained from New England Bio Labs, though some were from Miles or from Bethesda Research Laboratories. Fnu4H I was generously provided by Dr. David Russell. The large fragment of E. coli DNA polymerase I (Klenow fragment) was obtained from either Boehringer-Mannheim or New England Bio Labs. Dialysis tubing was purchased from Fisher Chemical Co. (size A). E. coli strain SF-8 (114) was used to grow plasmids.

Methods

B.1. Preparation of plasmid DNA

DNA of the plasmids pBR 322, pDt 17R, pDt 27R, and pDt 73 was obtained by a two-step procedure involving preparation of a cleared lysate and chromatography of the pancreatic RNase-treated, protease K-treated, phenol-extracted lysate on an agarose A-150 M column (45 cm x 1.4 cm), and provided for use by Dr. Allen Delaney. DNA of the plasmid pDt 16, prepared as described (114), was provided by D. Taylor and Dr. R.C. Miller, Jr.

B.2. Agarose gel electrophoresis

About 200 ml of agarose solution is required to fill the gel trays (22 cm x 15 cm x 0.6 cm) commonly used in our laboratory for agarose gel electrophoresis. Agarose solutions (0.5-1.4%, w/v) were prepared by dispersing the desired amount of agarose in 200 ml of 40 mM Tris (base), 20 mM NaH_2PO_4, 1 mM EDTA containing 1 µg/ml ethidium bromide and heating on a hot plate at about 95° with continuous stirring. Sample wells were formed by cooling the agarose solution with slot formers (1 cm x 1 cm x 0.12 cm) in place about 5 cm from one end of the gel tray. Also, two layers of Whatman 3MM filter paper were inserted as wicks into each end of the gel tray before pouring the gel. The agarose solution was allowed to gel for at least two hours before use. The electrophoresis buffer was the same as in the gel (40 mM Tris base, 20 mM NaH_2PO_4, 1 mM EDTA). Before loading the samples for electrophoresis, the gel surface was covered with this buffer and the wicks (placed in the electrode buffer compartments) thoroughly wetted. Samples (in 10% sucrose) were loaded into the slots using automatic pipettors with disposable tips. Usually, sample volumes were about 20 µl

per slot containing at least 1 μg of DNA. For preparative agarose gel electrophoresis, slots were 1 cm x 1 cm x 0.45 cm, and volumes loaded per slot were up to 80 μl. Prior to electrophoresis, the loaded, submerged gel was covered with Saranwrap.

Electrophoresis was carried out on a flat-bed Savant apparatus with water-cooling. Gels were generally run overnight (about 16 hr.) at 200V, though in some cases they were run at up to 400V without apparent loss of resolution.

DNA fragments in gels were visualized by illuminating with ultraviolet light. When desired, fragment patterns were photographed using a Polaroid camera (orange lens filter) and polaroid type 57 high speed film, illuminating the gel with short wavelength ultraviolet light.

A variety of different agarose concentrations was used in different experiments. Specific details concerning gel concentrations and DNA fragment mobilities are presented in individual figure legends.

B.3. Polyacrylamide gel electrophoresis of double-stranded DNA fragments

Native double-stranded DNA fragments were often prepared by electrophoresis on non-denaturing polyacrylamide gels. For fractionation of fragments 50 b.p. to 600 b.p. in length (or considerably more in some cases), gels (40 cm x 20 cm x 0.05 cm) containing 5% acrylamide, 0.25% methylene-bis-acrylamide, 90 mM Tris-borate (pH 8.3), 2.5 mM EDTA were used. These were polymerized by adding ammonium persulfate and TEMED to final concentrations of 0.05% (w/v) and 0.1% (v/v), respectively. However, polyacrylamide gels with concentrations as low as 2.25% acrylamide, 0.11% methylene-bis-

acrylamide were used to fractionate DNA fragments as large as 4.4 kb. To assure discrete DNA fragment bands, gels were run at low voltages (100-125 V) at room temperature.

^{32}P-labelled DNA fragments were located by autoradiography. Unlabelled DNA fragments were located by staining the gel 5 min. in a 1 μg/ml ethidium bromide solution, then illuminating the gel with ultraviolet light. Gel bands containing DNA fragments to be characterized further were excised with a scalpel and the DNA recovered by electroelution (below) or by soaking in 0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1%(w/v) SDS. Recoveries of most fragments up to 500 b p. long were 80-95% after soaking overnight at 37°. For longer DNA fragments, elution was for a longer time or at a higher temperature (up to 65°). Fragments were collected by ethanol precipitation.

B.4. Digestion of DNA with restriction enzymes

A number of different restriction enzymes were used to digest DNA. Reaction mixes contained DNA, buffer, the restriction enzyme(s), and sometimes bovine serum albumin to stabilize the enzyme(s). Reaction conditions were similar, though not always identical, to those recommended by commercial suppliers. For convenience, a limited number of standard buffers were used. (These were prepared at concentrations 10-fold higher than the desired concentration in the reaction mix and stored at -20°.) All reaction mixes contained 6 mM Tris-HCl (pH 7.5), 6 mM $MgCl_2$, and 6 mM 2-mercaptoethanol, plus 150 mM NaCl (for use with Xho I, Xba I, EcoRI), or 60 mM NaCl (for use with Sau96 I, Bgl II, Alu I, Pvu II, Hha I, Pst I, Hinf I, Sau3A I, Dde I),

or 6 mM NaCl (for use with Hae III, Taq I, Hind III, Ava I, Cla I, Xma I, BamH I, Fnu4H I), or 60 mM KCl (for use with Bcl I), or 6 mM KCl (for use with Msp I). Reactions were at 37° except with Taq I (used at 65°). DNA concentrations were estimated by measuring the A_{260} of DNA samples (50 µg/ml/A_{260}, either in H_2O or buffered at about pH 7). Reactions were incubated for at least twice the time estimated to be necessary for 100% digestion of the input DNA based on specified enzyme activity values (units/ml) and the volume of enzyme used. Bovine serum albumin was generally included in the reaction mixes (final concentration, 100 µg/ml) when reactions were for relatively long time periods (4 hr. or more).

The DNA concentration in digestion reaction mixes was 1 µg/µl or less. As contaminants present in some DNA samples inhibit the restriction enzymes when present at relatively high concentrations, the DNA concentration (and that of possible contaminants) was kept low. If a more concentrated DNA sample was required for subsequent steps (particularly for gel electrophoresis), the digested DNA was collected by ethanol precipitation. Sodium acetate was added to the reaction mix to a concentration of 0.3M, the DNA precipitated by adding 2.5 volumes of 95% ethanol and chilling (at least 1 hr. at -20°, or 20 min. at -70°) and collected by centrifugation in an Eppendorf high-speed table-top centrifuge (4 min. at room temperature). After discarding the supernatant liquid and drying briefly in a dissicator, the sample was ready to be dissolved and used further.

B.5. Electroelution of DNA fragments from gels

The portion of an agarose or polyacrylamide gel containing a desired DNA fragment was excised with a scalpel and placed in a glass

wool-plugged Pasteur pipet with a cut-off, fire-polished tip. Dialysis tubing (prepared by boiling four times for 5-10 min. in 1 mM EDTA, pH 7.0) was tied and clamped at one end, filled with 0.30 ml of electroelution buffer (20 mM Tris-acetate, pH 7.0) and slid over the narrow, plugged end of the pipet. The pipet was filled with the same buffer, and the tubing adjusted to remove any air bubbles at the pipet tip that would impede the flow of current. After placing the pipet in the electrophoresis apparatus, the bottom (anode) buffer chamber was filled to a level sufficient to complete an electrical circuit but below the top of the dialysis tubing on the pipet. The upper (cathode) buffer chamber was filled, and electrophoresis of DNA fragments from the gel slice to the dialysis tubing carried out at 150V (2 mA/tube) for 4-8 hr. at room temperature. (Recovery of [^{32}P] DNA can be checked by monitoring the dialysis tubing with a Geiger counter, while unlabelled, ethidium-stained DNA is detected by the presence of the orange dye in the dialysis tubing.) The DNA-containing eluate was removed from the dialysis tubing into a plastic test tube by low speed centrifugation, one-ninth volume of 3M sodium acetate added, and the DNA collected by ethanol precipitation and centrifugation as described above (section B.4.).

B.6. 3'-end labelling of restriction fragments

Restriction fragments with 5'-extended sequences were labelled at 3'-ends by repair synthesis on the 5'-extended template, using *E. coli* DNA polymerase I (Klenow fragment) and [α-^{32}P] labelled deoxynucleoside triphosphates (152).

Because of the high specificity of the enzyme, the repair reaction is very "clean" and it can be manipulated in many cases so that only one of two fragment ends (generated by two or at times by one restriction enzyme) is radiolabelled. This is accomplished by using an appropriate combination of unlabelled dNTP's plus a different [α-^{32}P]dNTP which is complementary to only one of two 5'-extended template sequences. In other cases, both ends of a restriction fragment are labelled, then separated by cleavage with a second restriction enzyme. (See Chapter IV, section A.3, for specific examples.)

Efficient labelling of restriction fragments by E.coli DNA polymerase I (0.5 - 1 unit) was obtained when [α-^{32}P] dNTPs (≥ 2000 Ci/mmole) were added to restriction enzyme reaction mixes (from 10 µl to 100 µl) in varying amounts (1-100 Ci; ≥ 0.05 Ci/µl). One µl of each appropriate unlabelled dNTP (0.5 mM) was added to the reaction mixes. Labelling reaction mixes contained 20-60 mM NaCl, 6 mM Tris-HCl (pH 7.5), 6 mM $MgCl_2$, and 6 mM 2-mercaptoethanol, as well as deoxynucleoside triphosphates, DNA, and Klenow polymerase. Reactions were at room temperature for 5-10 min.

Because of the high specific radioactivity and low concentration of the [α-^{32}P] dNTP's, only a fraction of the DNA ends present were usually radiolabelled. To ensure uniformity of the product where more than one radiolabelled product could occur by partial repair of an end (i.e. where it was possible to add one or more than one [^{32}P]-nucleotide to an end), a "cold chase" step was added to fill out the end completely. This was done by adding a large molar excess of the appropriate unlabelled dNTP (2 µl; 0.5 mM). The reaction was continued for 2-3 min. at room temperature, then stopped by adding one-fourth volume of 60% sucrose, 10 mM EDTA, 0.05 % XC and BB, and chilling in an ice-water bath.

Better results were obtained using Klenow polymerase from Boehringer-Mannheim. The enzyme preparation obtained from New England BioLabs was found to contain a significant amount of contaminating 3'-exonuclease activity, which resulted in high backgrounds when apparently uniform fragments obtained as discrete bands by polyacrylamide gel electrophoresis were subjected to Maxam-Gilbert sequence analysis (below).

B.7. Size determinations for DNA fragments

Sizes of DNA fragments were estimated based on electrophoretic mobilities relative to standards of known length. This was done either graphically, establishing a standard curve by plotting inverse mobilities as a function of the known sizes of standard DNA fragments; or using a linear regression computer program (written by Allen Delaney, UBC) based on ref. 147. Standards used were HinfI-cut pBR 322 (for the size range 75-516 b.p.; ref. 148) or HindIII-cut bacteriophage λDNA (for the size range 0.54 kb to 23 kb; ref. 149). When ^{32}P-labelled DNA standards were desired, a small amount (~1μg) HinfI-cut pBR 322 or HindIII-cut λ DNA was 3'-end labelled with [α-^{32}P]dATP (1-5 μCi; 2000 Ci/mmole) and E. coli DNA polymerase I (Klenow fragment), then mixed with 5-10 μg of unlabelled DNA (cut with the same enzyme) in 10% sucrose, 2.5 mM EDTA, 0.02% (w/v) XC and BB.

B.8. Restriction enzyme mapping of recombinant plasmids

The recombinant plasmids used in the work presented here contain HindIII fragments from D. melanogaster genomic DNA inserted into the single HindIII site of the plasmid vector pBR 322 (114). Restriction sites in the plasmid DNA are known, as the complete nucleotide sequence of pBR 322 has

been determined (148). The unique array of restriction sites in the insert DNA was determined as follows. Restriction mapping involved two steps. First, a number of different restriction enzymes recognizing hexanucleotide sequences were used to digest recombinant plasmids. The resulting fragments (or uncut plasmid) were separated according to size by agarose gel electrophoresis, and their sizes determined relative to DNA fragments of known size run on the same gel (see Section B.7., above). Since the locations of restriction sites in pBR 322 DNA were known, enzymes cutting in the insert were identified, and in some cases cleavage sites could be precisely located in the insert, at this stage. Second, combinations of two enzymes found to cut in the insert were used to digest recombinant plasmid DNA, the digestion products separated by agarose gel electrophoresis, and fragment sizes determined as above. The sizes of all fragments seen on agarose gels were summed in each case to ensure against missing one of two fragments that co-migrate on the gels.

B.9. Identification of DNA fragments containing $tRNA^{Ser}$ sequences. Southern blotting experiments (150)

B.9.a. Preparation of [3'-^{32}P]$tRNA_7^{Ser}$ as a hybridization probe

Purified $tRNA_7^{Ser}$ was 3'-end labelled with *pCp and RNA ligase as described in section A.3 of this chapter. [^{32}P]$tRNA_7^{Ser}$ was separated from *pCp, [γ-^{32}P]ATP, and $Na_2H^{32}PO_4$ in the labelling reaction mix by Sephadex G-25 column chromatography. The sample was eluted from the column (13 cm x 0.7 cm) in 2 x SSC (0.30 M NaCl, 0.03 M sodium citrate), collecting 5 drop fractions manually in 1.9 ml polypropylene tubes. The ^{32}P-labelled

tRNA and starting materials were located by monitoring each fraction with a Geiger counter. [^{32}P] tRNA separated cleanly from radiolabelled starting materials in this way. Depending on the specific radioactivity of *pCp, up to 6 x 10^6 Cerenkov cpm (30-40% of input counts) were recovered in the first (tRNA) peak of radioactivity.

B.9.b. Southern blots from agarose gels

DNA fragments were digested with restriction enzymes and separated according to size by agarose gel electrophoresis, then denatured, transferred to a nitrocellulose membrane and hybridized with a [^{32}P]tRNA$_7^{Ser}$ probe, in a procedure based on that of Southern (150). The section of the agarose gel containing the origin and DNA fragments was excised with a scalpel and transferred from the gel tray to a glass tray. DNA in the gel was denatured by immersing the gel in 0.5 M NaOH, 1.5 M NaCl for 5 min. at room temperature. The NaOH/NaCl was removed and replaced with 3 M NaCl, 0.5 M Tris HCl (pH 7.0) to neutralize remaining NaOH (10 min. at room temperature). The gel was placed on three layers of Whatman 3MM filter paper (thoroughly wetted with 3M NaCl, 0.5 M Tris HCl (pH 7.0), then covered with a strip of nitrocellulose membrane (cut to the dimensions of the gel section) that had been uniformly wetted in the same buffer. Special care was taken at this step to ensure that no air bubbles were trapped between the nitrocellulose strip and the gel. The nitrocellulose strip was covered first with two layers of Whatman 3MM paper, then with absorbent paper towels. A glass plate was placed on top of the assembly (to maintain good contact throughout for efficient wicking) and the edges were covered closely with

polyvinyl chloride film to minimize evaporation. Transfer was for 4 hr. at room temperature. As monitored by [^{32}P] size standards present in some gels, transfer efficiency was at least 90%. Transferred DNA was cross-linked to the nitrocellulose membrane strip by baking for 2 hr. at 80° (on a glass plate to prevent curling). For hybridizations, the nitrocellulose strip with bound DNA was wetted on the glass plate with [^{32}P] $tRNA_7^{Ser}$ (at least 4000 Cerenkov cpm/cm^2 of membrane) diluted with 2 x SSC. The membrane strip was covered with mylar film, a second glass plate, the assembly tightly wrapped with polyvinyl chloride film and clamped together. Hybridization was at 65° for 6 hr. Following hybridization, the nitrocellulose strip was washed 4-5 times in 2 x SSC to lower radioactive background, dried on a glass plate, and DNA bands hybridizing [^{32}P] $tRNA_7^{Ser}$ located by autoradiography (2-4 days).

B.9.c. Southern blots from polyacrylamide gels

Restriction fragments were separated according to size on non-denaturing polyacrylamide gels (section B.3.). DNA in the gel, on a backing glass plate, was denatured and neutralized as for agarose gels (above). Transfer of DNA to nitrocellulose was accomplished as follows. A nitrocellulose sheet (cut to the dimensions of the gel) was placed on the neutralized polyacrylamide gel (avoiding trapping air bubbles), then removed from the glass backing plate (together with the tightly adhering gel) by inverting the glass plate and peeling off the nitrocellulose sheet and gel. The gel, now on top, was covered with three layers of Whatman 3MM paper (thoroughly wetted in 3M NaCl, 0.5 M Tris HCl, pH 7.0), then inverted

again so that the nitrocellulose strip was again on top. Three layers of dry Whatman 3MM paper were placed on top to absorb liquid. The entire assembly was closely covered with polyvinyl chloride film, then a glass plate and lead weights were placed on top to maintain good contact throughout for uniform, efficient transfer. Transfer was for 16 hr. at room temperature. As monitored by [^{32}P] DNA size standards included in the gel, transfer was essentially quantitative. Filters were baked, hybridized with [^{32}P] $tRNA_7^{Ser}$, washed, and hybridizing DNA bands located by autoradiography as for agarose gel transfers (above).

B.10. Mapping of single-end labelled DNA fragments by partial digestion with restriction enzymes - Smith/Birnstiel mapping

Restriction mapping of single-end labelled DNA fragments was performed essentially as described by Smith and Birnstiel (151). A single-end labelled [^{32}P] DNA fragment was prepared by specifically labelling one 3'-end with an [α-^{32}P] dNTP and the large fragment of E. coli DNA polymerase I (section B.6., above). Aliquots of the purified, single-end labelled DNA fragment were partially digested with a number of restriction enzymes. Each digestion reaction (20 µl) contained [^{32}P] DNA (at least 20,000 Cerenkov cpm), 1 µg of sheared calf thymus DNA as carrier, 100 µg/ml bovine serum albumin, and restriction buffer (see "Digestion with restriction enzymes," above). Restriction enzyme (about 4 units, sufficient to digest 1 µg of DNA in 15 min.) was added to start the reaction, and 5 µl aliquots were removed at 2, 4, 8, and 12 min. and added to 1 µl of 200 mM EDTA (pH 4.7) in a polypropylene tube chilled on solid carbon dioxide. (This amount of EDTA is in 1.7-fold molar excess over Mg^{+2} in the reaction mix.) The

pooled, frozen sample was dried over P_2O_5 in a dessicator under aspirator vacuum, redissolved in 20% (w/v) sucrose, 5 mM EDTA, 0.04% (W/v) XC and BB (10 µl per gel slot to be run), and electrophoresed on a non-denaturing polyacrylamide gel. Radiolabelled restriction fragments were located by autoradiography. The distances of various restriction sites from the radiolabelled end were determined by reference to the mobilities of ^{32}P-labelled DNA fragments of known size run on the same gel (see section B.7., above).

B.11. Sequence analysis by the method of Maxam and Gilbert (132)

Samples of single-end labelled DNA fragments (that were obtained from about 20 µg plasmid, and contained at least 50,000 Cerenkov ^{32}P-cpm in 30 µl H_2O) were sequenced by the chemical modification method of Maxam and Gilbert (132). The protocol used in our lab is closely related to the one described by those authors (an excellent description of the method is found in ref. 152, to which the reader is directed for detailed information).

Four modification reactions were used, specific for G, A+G, C+T, or C. Reactions were carried out in conical 1.9 ml polypropylene tubes with tightly closing caps.

Modification of G with dimethylsulfate (DMS)

The procedure was carried out in a fume hood. The reaction mix contained one-sixth of the [^{32}P] DNA sample, 1 µg calf thymus DNA, 50 mM sodium cacodylate (pH 8.0), 10 mM $MgCl_2$, and 1 mM EDTA (200 µl; room temperature). The reaction (3 min. at room temperature) was started by adding 2 µl of DMS with a pipettor (disposable tips). It was stopped by adding

50 µl of "stop mix" (1.5 M sodium acetate, pH 7.0, 1 M 2-mercaptoethanol, 100 µg/ml E. coli tRNA; 0°) then 750 µl 95% ethanol (-70°; chilled 30-60 min. on solid CO_2) and mixed by inverting four times. After storing for 20-30 min. at -70° on solid CO_2, the DNA sample was collected by centrifugation (4 min. at room temperature, Eppendorf table-top centrifuge). The supernantant liquid was carefully removed with a Pasteur pipet, and mixed with 5M NaOH to destroy the DMS. Pipettor tips contaminated with DMS were treated in the same way. The pellet was dissolved in 200 µl of 0.3 M sodium acetate by agitating it vigorously on a Vortex mixer, then the DNA was precipitated with 500 µl 95% ethanol (-70°) and collected by centrifugation as described above. The supernatant liquid was discarded without special precaution. The pellet was finally washed with 500 µl 95% ethanol (-70°), centrifuging and discarding the supernatant liquid as in the previous step. The pellet was then dried in a dessicator for 5-10 min. under aspirator vacuum. Strand scission of modified DNA was accomplished by dissolving the pellet in 100 µl of 10% piperidine (freshly diluted) and heating for 40 min. at 90° in the tightly closed 1.9 ml tube. After this step. the tubes were cooled and condensate collected by centrifuging briefly (1 sec., Eppendorf). The samples were dried overnight from the open tubes in a dessicator containing P_2O_5 and dried molecular sieves, under aspirator vacuum. Samples were then dissolved in 25 µl H_2O and dried in an oven (45°) with forced air circulation. This procedure was repeated three times. (At this point, there should be no residual smell of piperidine.)

Finally, samples were dissolved in 45 mM Tris-borate (pH 8.3), 1.25 mM EDTA, 7M urea, 0.05% XC and BB (5 µl per sequencing gel to be run), heated 1 min. at 100°, chilled on ice, then applied (with a calibrated, drawn-out capillary tube) to thin, denaturing polyacrylamide gels (ref. 135;

prepared as described in section A.7. above) and electrophoresed. Gels were run at constant power (20 watts) until xylene cyanol had migrated 40% the length of a 20% gel, 110% the length of a 12% gel, 135-150% the length of an 8% gel, or 100-150% the length of a 6% gel. Gel bands were located by autoradiography (see figure legends in Chapter IV for specific details regarding ^{32}P-cpm loaded per slot and time of autoradiographic exposure in individual experiments).

The A+G reaction: depurination in formic acid

The reaction mix (20 µl) contained one-third of the [^{32}P]DNA sample and 1 µg of calf thymus DNA, at room temperature. The reaction was started by adding 3 µl of 10% formic acid (0°) and placing the 1.9 ml tube in a water bath at 37°. After 13 min. at 37°, the reaction was stopped by adding 230 µl of "stop mix" (0.3M sodium acetate, 0.1 mM EDTA, 20 µg/ml E. coli tRNA; 0°), then 750 µl 95% ethanol (-70°), and mixed by inverting four times. Ethanol precipitations and all subsequent steps were performed as described for the G reaction, above.

Modification of C+T with hydrazine

The procedure was performed in a fume hood. To the reaction mix (20 µl, at room temperature), containing one-third of the [^{32}P]DNA sample and 1 µg calf thymus DNA, was added 30 µl of hydrazine (room temperature).

After 10 min. at room temperature, the reaction was stopped by adding 200 µl of "stop mix" (0.3M sodium acetate, 0.1 mM EDTA, 20 µg/ml E. coli tRNA; 0°), then 750 µl 95% ethanol (-70°) and mixing by inverting

four times. (After centrifugation, the supernatant liquid, as well as any pipettor tips contaminated with hydrazine, was discarded into 2M $FeCl_3$. Ethanol precipitations and all subsequent steps were otherwise the same as described above for the G reaction).

Modification of C with hydrazine

The procedure was similar to that just described but at a high NaCl concentration. To the reaction mix (20 µl, at room temperature) containing one-sixth of the [^{32}P] DNA sample, 1 µg calf thymus DNA, and 3.75 M NaCl, was added 30 µl of hydrazine (room temperature). After 14 min. at room temperature, the reaction was terminated by adding 200 µl "stop mix." Then 750 µl 95% ethanol (-70°) was added and mixed by inverting the tube four times. The product was collected by centrifugation and treated as described above.

By the procedure described, it was possible to obtain a continuous sequence of up to 240 nucleotides on the labelled DNA strand.

Chapter III

SEQUENCE ANALYSIS OF SERINE tRNAs IN DROSOPHILA

A. Introduction

The standard genetic code contains six codons that are translated to serine in polypeptides (UCN, AGC, and AGU). One might expect a family of serine isoacceptor tRNAs to be complex relative to other isoacceptor families reading fewer codons. Certainly this is true in Drosophila melanogaster, where there are five major and two minor $tRNA^{Ser}$ species, making this isoacceptor family more complex than all but one or two others in the organism (91).

Serine tRNAs in Drosophila melanogaster were the subject of a study by White et al. (102) in which several isoacceptors were purified and their nucleotide contents and coding specificities examined. These workers reported that the five major $tRNA^{Ser}$ species could be divided into two groups based on similar nucleotide content. The first group consisted of $tRNA_2^{Ser}$ and $tRNA_5^{Ser}$. These tRNAs were found to contain t^6A or mt^6A and to respond to AGC and AGU triplets in ribosome-binding assays. However, $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$, which contain i^6A rather than t^6A or its derivatives, responded to UCG and UCU triplets, respectively. Though the coding properties of $tRNA_6^{Ser}$ (the other major serine isoacceptor) were not determined, its nucleotide content (including i^6A) was found to be very similar to those of

$tRNA_4^{Ser}$ and $tRNA_7^{Ser}$; thus, it also probably reads codons of the UCN group. The two groups of serine tRNAs are apparently structurally distinct, though functionally identical except for codon specificities.

Genes for a purified Drosophila isoacceptor tRNA can be localized on polytene chromosomes from salivary glands of the insect by hybridization of the tRNA in situ (see Chapter I, section D.2.a.). The banding patterns of such chromosomes are known, and a composite photographic map has been assembled (153). The cytogenetic map has been correlated with the genetic map of Drosophila, established by analysis of recombination frequencies, in many cases (154).

Genes for $tRNA_{2b}^{Ser}$ (155), $tRNA_4^{Ser}$, and $tRNA_7^{Ser}$ (109) have been localized by in situ hybridization. Genes for $tRNA_{2b}^{Ser}$ are found at several regions of chromosome 3: 86A, 88A9-12, and 94A6-8 (155). These regions are all distinct from the hybridization sites for $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$. The latter two tRNAs both hybridize to a major site at 12DE on the X chromoxome, and to minor sites at 23E on the left arm of chromosome 2 (2L), 56D on chromosome 2R, and 64D on chromosome 3L. They are indistinguishable by RNA-DNA hybridization in situ, as either tRNA competes with the other for hybridization (109,155).

These observations raised certain questions regarding $tRNA_4^{Ser}$, $tRNA_7^{Ser}$, and the genes coding for these isoacceptors, most notably: do the two isoacceptors arise from the same genes and differ only in extent of modification; or do they arise from similar but distinct genes? It was to answer this initial question that sequence analysis of $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ was undertaken.

Other reasons for analyzing the sequences of serine tRNAs from Drosophila can be envisaged. By comparing sequences of several isoacceptor species, we might gain some insights into how tRNA genes evolve. Also, by sequence comparison of related tRNAs it might be possible in some cases to identify structural features recognized by the cognate aminoacyl-tRNA synthetase. An important consideration in undertaking this analysis was to establish a well-characterized system for studying tRNA gene expression. Transfer RNAs interact with many cell constituents. Such interactions potentially can be studied genetically, and Drosophila melanogaster is one of a very few multicellular organisms amenable to both extensive genetic and biochemical analysis at this time. As a model tRNA system, serine tRNAs have certain desirable features. For example, nonsense suppressor tRNAs can be generated by a single base change in structural genes for $tRNAs^{Ser}$ responding to UCN codons. Also, it may prove possible to isolate mutant Drosophila defective for tRNA modifying enzymes. An enzyme that introduced a distinctive modification into only a few tRNA species - for example, the enzymes involved in synthesis of 4-acetylcytidine, 3-methylcytidine, or 6-isopentenyladenosine (modifications found in yeast, Drosophila, and rat liver serine tRNAs) - would be easier to identify than an enzyme modifying many tRNAs. In any case, identification of such mutant enzymes would require that the sequence of the relevant wild type tRNA be known, including modifications. Finally, $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ were of interest to us because the major site of in situ hybridization is at 12DE on the X chromosome. Genes for about 25 purified D. melangoaster tRNAs have been localized. Of these tRNAs, only $tRNAs_{4,7}^{Ser}$ hybridize reliably to a site on the X chromosome.

(Under some hybridization conditions, $tRNA_5^{Lys}$ also goes to the 12DE region.) Isolation of mutants deficient for parts of the 12DE region seems to be a possible approach to studying in vivo expression of tRNA genes. This approach has been used previously to study $tRNA^{Val}$ genes on chromosome 3 (113). In principle, isolation of X-linked mutations is an easier task.

Dunn et al. reported the isolation of five distinct recombinant plasmids hybridizing $tRNAs_{4,7}^{Ser}$ (114). Four of the plasmids (pDt16, 17R, pDt 27R, and pDt 73) derive from the 12DE region while the fifth plasmid (pDt 1, pDt 5, and pDt 81 contain the same insert) is from 23E on chromosome 2L (114,155,156). Studies on expression of $tRNA^{Ser}$ genes from sex chromosomes and autosomes would be of interest. Male flies contain one functional X chromosome and females two X chromosomes, since there is no Barr body formation in Drosophila as is found in mammals. Thus, it may be that expression of X-linked and autosomal $tRNA^{Ser}$ genes is controlled differently. Preliminary to studies on expression of $tRNA^{Ser}$ genes in recombinant plasmids, characterization of the genes as well as of the tRNA products would be needed.

Serine tRNAs and their genes in Drosophila melanogaster constitute a system that could prove useful in a wide variety of genetic, molecular biological, and biochemical studies. This thesis presents initial characterization of this system. The sequences of $tRNA_4^{Ser}$, $tRNA_7^{Ser}$, and $tRNA_{2b}^{Ser}$ have been determined; the sequence analyses are presented here in Chapter III. In Chapter IV I present data concerning the structures and organization of $tRNA^{Ser}$ genes located on the X chromosome of Drosophila melanogaster.

B. Results

The nucleotide sequences of serine tRNAs from Drosophila melanogaster were determined by a combination of rapid sequencing techniques. The strategy employed consisted of three steps.

(i) The nucleotide content of the tRNA was analyzed.

(ii) The nucleotide sequence was analyzed by the method of Stanley and Vassilenko. Nucleotides were analyzed by thin layer chromatography as [5'-^{32}P] nucleoside 5', 3'-bisphosphates (*pNp's), and where there was reason to believe the base was modified, as nucleoside [5'-^{32}P] phosphates (*pN's) in two solvent systems.

(iii) The nucleotide sequences of terminally radiolabelled tRNAs or oligonucleotides were analyzed by the gel read-off method. In every case, both 5'- and 3'-end labelled tRNAs were analyzed.

In some cases, the above analyses were supplemented by two dimensional "wandering spot" analysis of end-labelled RNAs. By such combinations of methods, direct information on every nucleotide in each tRNA sequence was obtained. Each of the methods listed above is described in Section A of Chapter II.

The sequence analysis of each $tRNA^{Ser}$ is described separately. Analysis of $tRNA_7^{Ser}$ is presented first, in considerable detail. Analyses of $tRNA_4^{Ser}$ and $tRNA_{2b}^{Ser}$ were similar, and are therefore discussed less extensively.

B.1. Sequence analysis of D. melanogaster $tRNA_7^{Ser}$

Every nucleotide of the $tRNA_7^{Ser}$ sequence could be identified primarily on the basis of direct evidence obtained from gel "read-off", "wandering spot", or Stanley/Vassilenko experiments (Chapter II). Interpretation of sequence data was at times strengthened by indirect arguments based on homology with the equivalent tRNA in other eukaryotes and on base-pairing constraints. In general, each nucleotide of the sequence was identified unambiguously in independent experiments, usually by two different methods.

Nucleotide analysis indicated the presence in $tRNA_7^{Ser}$ of the standard nucleotides pA, pG, pC, and pU; and of the modified nucleotides pI, pG_m, pac^4C, pT, pψ, pD, pU_m, pm^5C, pm^3C, and pm^1A (Fig. 4a). Neither pm_2^2G nor pi^6A was seen, though both were found in $tRNA_7^{Ser}$ by White et al. (102). Both nucleotides were identified in the tRNA during Stanley/Vassilenko experiments, however. Their absences here are attributed to poor kinase-labelling of i^6Ap, and overlap of $*pm_2^2G$ with the large *pA spot. The modified nucleotides pI, pG_m, pac^4C and pU_m were identified by co-migration with commercially obtained nucleotide standards on two dimensional thin layer chromatography (Chapter II, A.8.g.). Others were identified by their chromatographic mobilities, based on published values (42). The relative amounts of the various nucleotides were quantitated by eluting the radioactive nucleotides from each spot and determining radioactivity as Cerenkov cpm. The number of each nucleotide present per $tRNA_7^{Ser}$ molecule was estimated relative to pT,which should be present once per tRNA. The results of such analysis are presented in Table 1.

Figure 4 - Nucelotide analysis of Drosophila serine tRNAs.

Nucleoside [5'-^{32}P] phosphates were prepared from (a) $tRNA_7^{Ser}$, (b) $tRNA_4^{Ser}$, and (c) $tRNA_{2b}^{Ser}$ as described in Chapter II. Samples of 3 µl were applied to each thin layer plate. After chromatography the thin layer plates were exposed to Kodak XR-1 X-ray film for (a) 6 hr., (b) 30 hr., and (c) 16 hr.

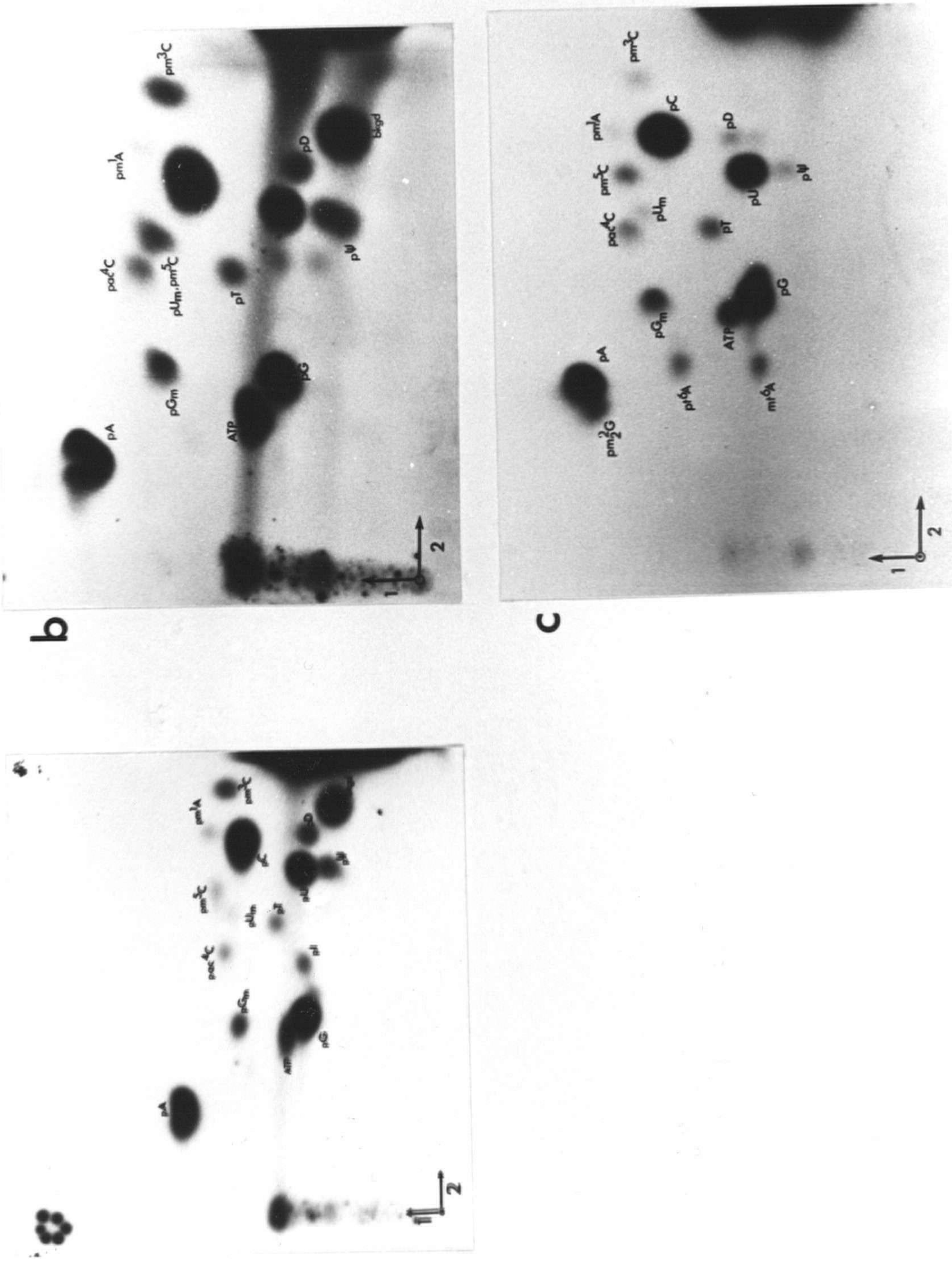

a
b
c
pA
pGm
ATP
pG
pT
pac4C
pm1A
pm3C
pm5C
pUm
pD
pψ
pC
pU
Bkgd
1
2

Table 1 - Nucleotide content of serine tRNAs

	$tRNA_7^{Ser}$	$tRNA_4^{Ser}$	$tRNA_{2b}^{Ser}$
pG	20.7[a] (21)[b]	16.7[a] (22)[b]	24.1[a] (22)[b]
pA	12.5 (13)	9.3 (13)	16.3 (12)
pC	17.7 (19)	16.9 (19)	21.2 (19)
pU	12.1 (10)	10.7 (10)	12.5 (11)
pm_2^2G	-- (1)	-- (1)	0.85 (1)
pG_m	1.6 (2)	1.2 (2)	0.86 (1)
pi^6A	-- -	-- -	-- -
pt^6A	-- -	-- -	1.85 -
pmt^6A	-- -	-- -	N.D.[c] (1)[c]
pm^1A	0.39 (1)	0.24 (1)	0.38 (1)
pI	1.2 (1)	-- -	-- -
pm^3C	1.45 (2)	1.55 (2)	1.55 (2)
pac^4C	0.68 (1)	0.69 (1)	0.72 (1)
pm^5C	0.70 (1)	1.5[d] (1)	1.25[d] (1)
pU_m	0.28 (1)	-- (1)	-- (1)
pT	=1.00 (1)	=1.00 (1)	=1.00 (1)
pD	1.79 (2)	2.0 (3)	0.96 (3)
pψ	3.2 (2)	2.6 (2)	1.7 (4)

a - Values expressed relative to pT, which is given a value 1.00.

b - Values in parentheses are the expected values based on the nucleotide sequences determined. They are adjusted to reflect the number of Np's expected on digestion of the tRNAs with ribonuclease T_2; thus the sum of these values is less than 85.

c - Value not determined. The sum of values for pt^6A and pmt^6A should be one.

d - This value is the sum of pm^5C and pU_m, which were not separated by chromatography.

Most of the $tRNA_7^{Ser}$ nucleotide sequence (positions 6-82), including modified nucleotides, was determined by the method of Stanley and Vassilenko. The ammonium sulfate solvent system used to separate pGp, pAp, pCp, and pUp on PEI cellulose plates at times allowed modified nucleotides to be distinguished (Fig. 5d). Further characterization of these nucleotides, or of others deemed likely because of their locations in the tRNA "cloverleaf" to be modified, was accomplished by digesting the same [^{32}P]-oligonucleotide with nuclease P_1 to generate the 5'-terminal *pN. Identification of modified nucleotides as *pN's was generally straightforward, as described above for nucleotide analysis. (See Figures 5f, 5h.)

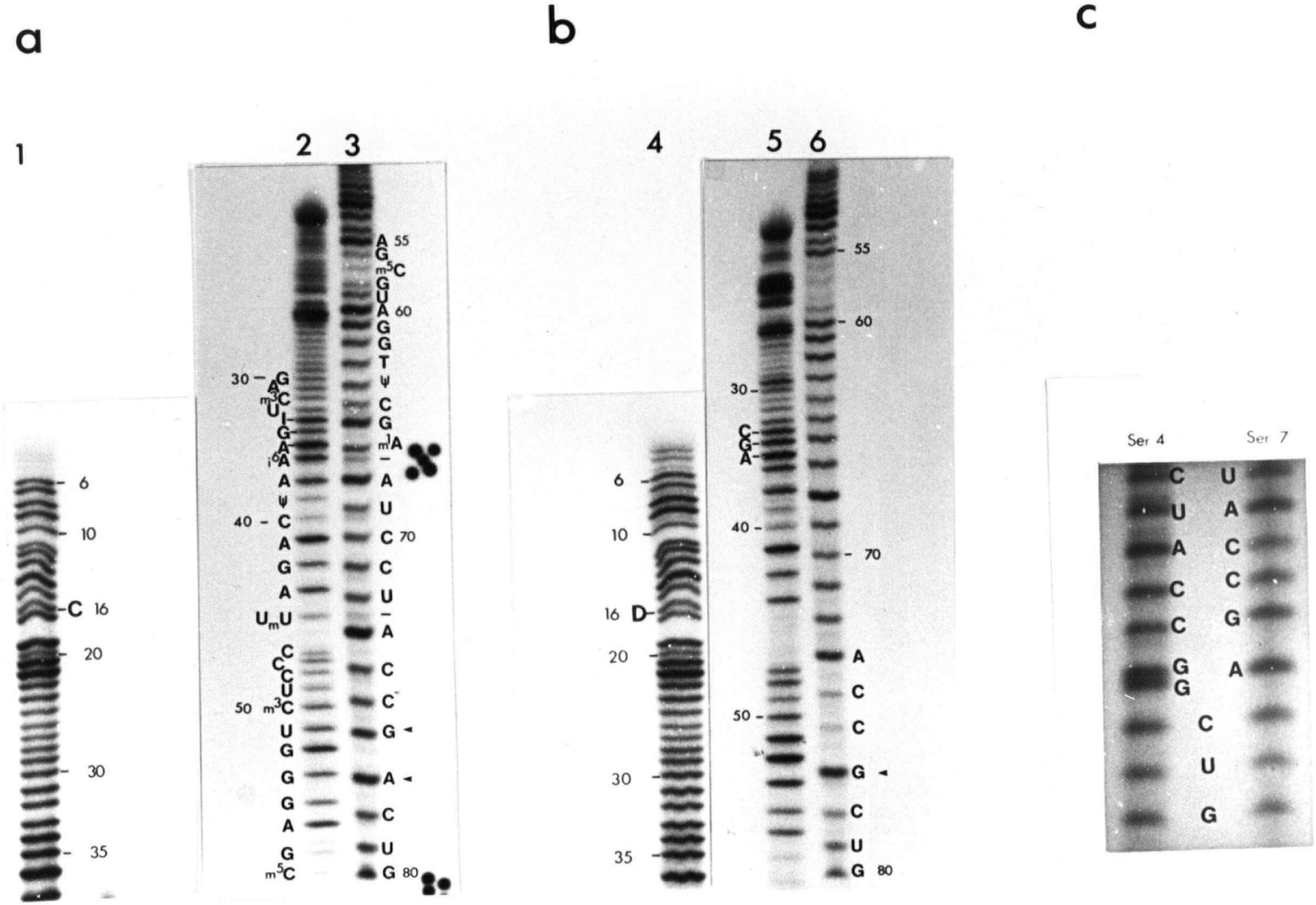
The entire nucleotide sequence of $tRNA_7^{Ser}$ was spanned in gel "read-off" analyses of the [5'-^{32}P]tRNA, [3'-^{32}P]tRNA, or a 5'-end labelled 3'-half molecule (including positions 36-85) that was fortuitously recovered in high yields from polynucleotide kinase labelling reactions. Much of the sequence could be determined directly from such experiments (see figures 6a, 6b, 6c, 6e, 6g). In conjunction with the results of Stanley/Vassilenko experiments, nearly all of the $tRNA_7^{Ser}$ sequence was thus determined unambiguously.

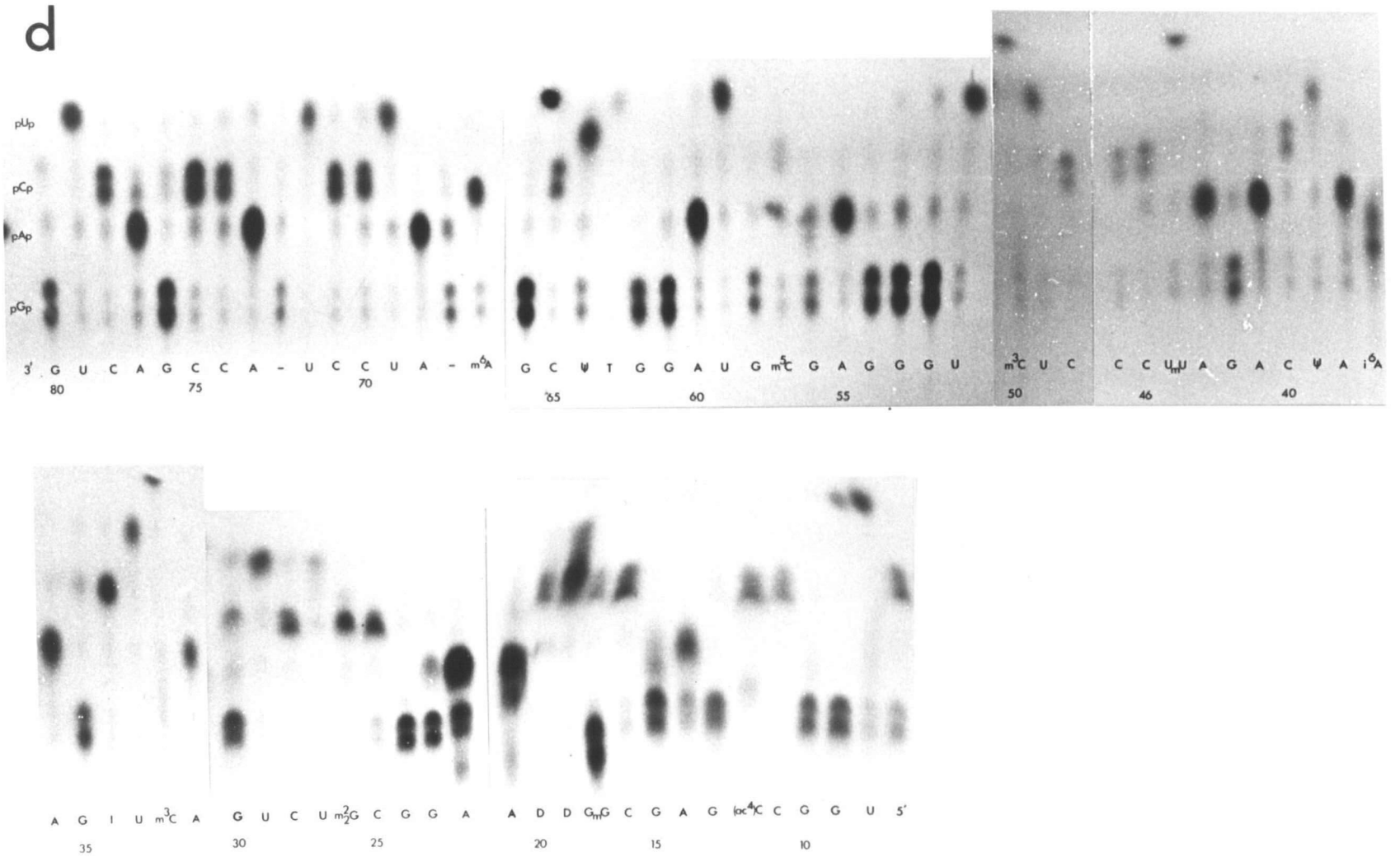
Analysis of the 5'-terminal sequence (positions 1-7) by a second method was deemed necessary to identify a 2'-O-methyl nucleotide at position 4 shown to be present in gel "read-off" experiments. Therefore, the sequence of nucleotides 1-12 was determined by the "wandering spot" method (two dimensional electrophoresis/homochromatography; see Chapter II).

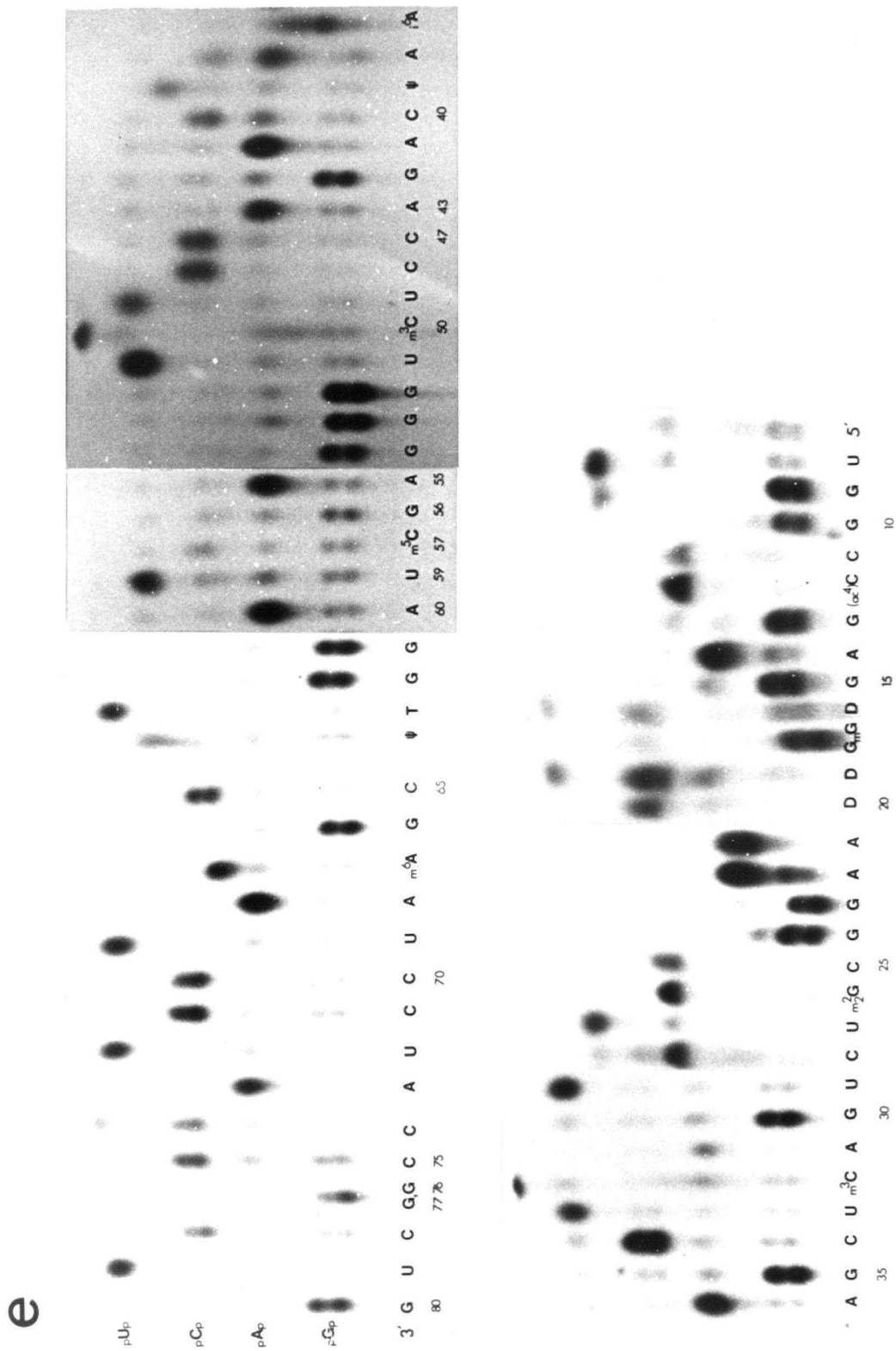
The only positions in the $tRNA_7^{Ser}$ sequence not determined by two independent methods were the universal 3'-terminal CCA. These nucleotides

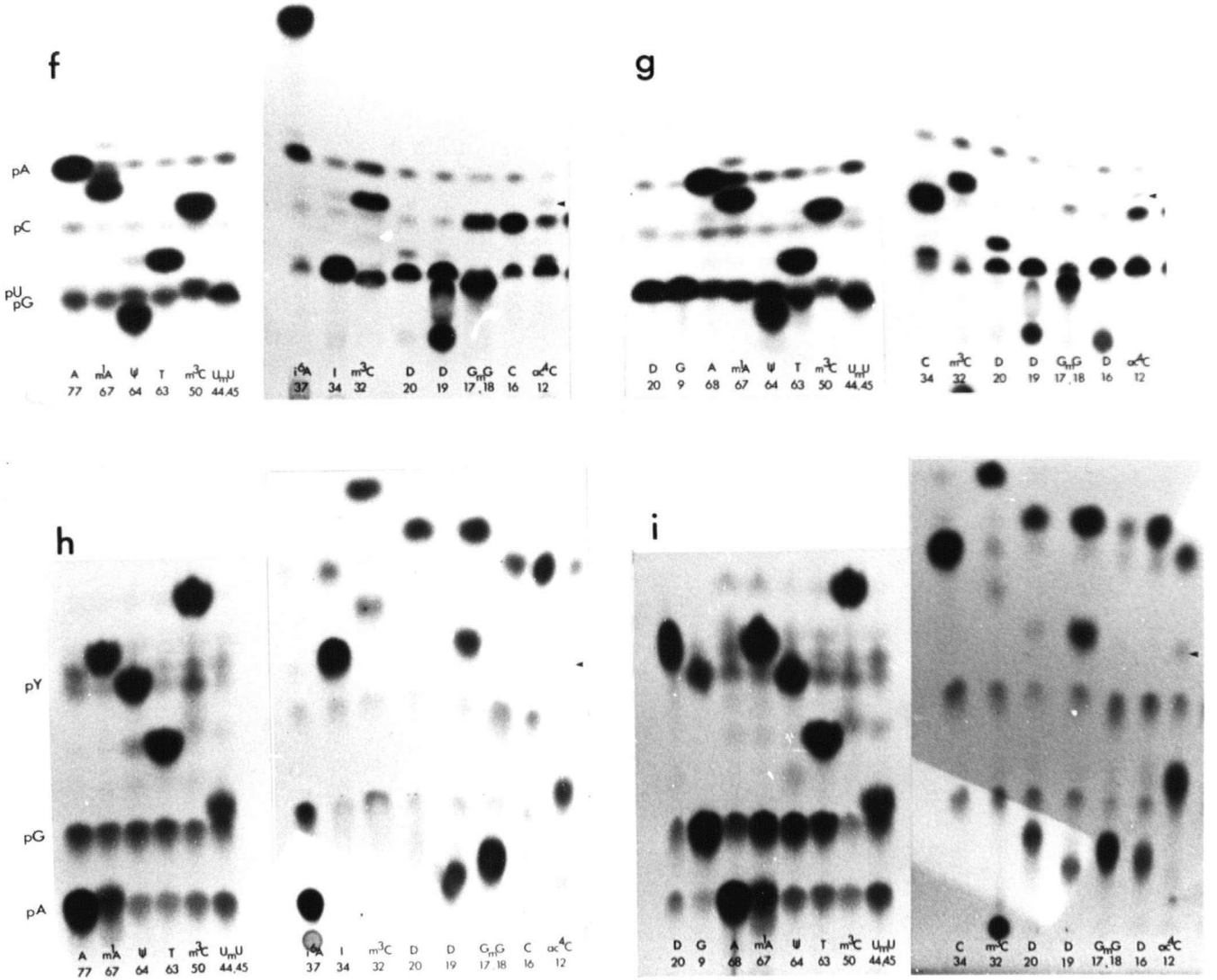
Figure 5 - Stanley/Vassilenko sequence analysis of $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$.

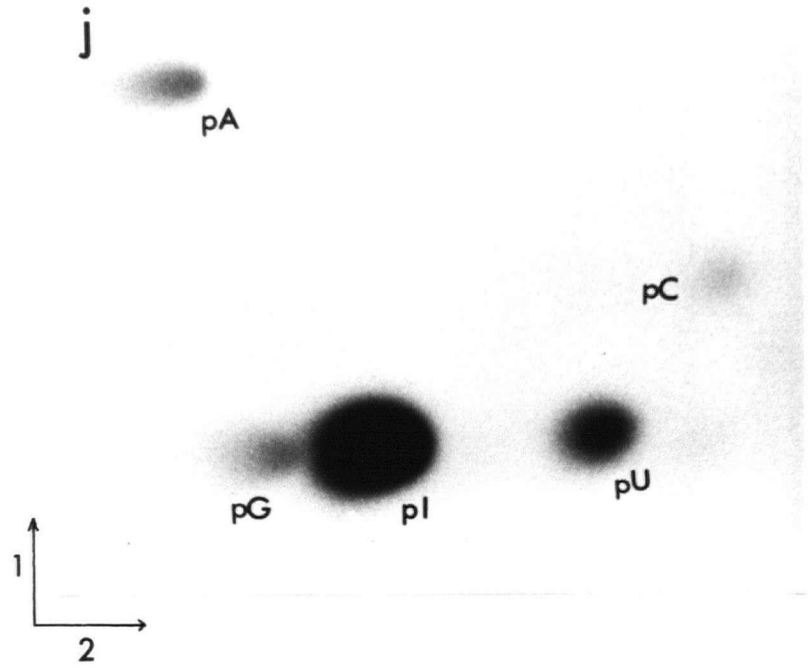
Hydrolysis of purified $tRNAs^{Ser}$ and 5'-end labelling were performed as described in Chapter II. Sections (a) and (b) show the results of the electrophoretic separation of [^{32}P] oligonucleotides from $tRNA_7^{Ser}$ and $tRNA_4^{Ser}$, respectively, on 20% polyacrylamide gels. Slots 1 and 4 contain 1.7 µg of $tRNA_7^{Ser}$ and 1.4 µg of $tRNA_4^{Ser}$ (8 µl/slot). The labelling reactions contained 37 µM [γ-^{32}P]ATP (1300 Ci/mmole). Electrophoresis was for 48 hr. at 1000V. Autoradiography was for 10 min. using "no-screen" X-ray film. Slots 2,3,5, and 6 each contain 1.2 µg of $tRNA^{Ser}$ (10 µl/slot). The labelling reactions contained 30 µM [γ-^{32}P]ATP (500 Ci/mmole). Electrophoresis was for 17 hr. at 1200V (slots 2,5) or 7 hr. at 1200V (slots 3.6). Autoradiography was for 2.5 hr. using "no-screen" X-ray film. The samples of [^{32}P] oligonucleotides from $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ shown in section (c) were prepared as described for slot 1 and 4 (above), and separated by electrophoresis for 8 hr. at 1000V. Autoradiography was for 11 min. using "no-screen" X-ray film. Sections (d) and (e) show the results of analysis of *pNp's from oligonucleotides derived from $tRNA_7^{Ser}$ and $tRNA_4^{Ser}$, respectively, as described in Chapter II. Samples contained at least 150 Cerenkov cpm, and were visualized by autoradiography for not more than four days ("no-screen" X-ray film). The results of cellulose thin layer chromatography of *pN's from $tRNA_7^{Ser}$ and $tRNA_4^{Ser}$ in solvent A are shown in sections (f) and (g), respectively. Results of chromatography of *pN's from $tRNA_7^{Ser}$ and $tRNA_4^{Ser}$ in solvent B are shown in sections (h) and (i), respectively. At least 250 Cerenkov cpm of each sample was chromatographed. Autoradiography was for 2 days using "no-screen" X-ray film. The results of thin layer chromatography of *pI 34 (6000 Cerenkov cpm) from $tRNA_7^{Ser}$ is shown in (j), while *pC 16 (2700 cpm) from $tRNA_7^{Ser}$ and *pD 16 (2900 cpm) from $tRNA_4^{Ser}$ are shown in (i). Autoradiography was for 3.5 days using "no-screen" X-ray film.


a
b
c
1
2
3
4
5
6
Ser 4
Ser 7


d
pUp
pCp
pAp
pGp
3' G U C A G C C A – U C C U A – m⁶A
80
75
70
G C Ψ T G G A U G m⁵C G A G G G U
65
60
55
m³C U C
50
C C U m U A G A C Ψ A i⁶A
46
40
A G I U m³C A
35
G U C U m₂²G C G G A
30
25
A D D G m G C G A G ac⁴C C G G U 5'
20
15
10


e
pUp
pCp
pAp
pGp
3' G U C G G C C A U C C C U A m6A G C Ψ T G G
80 77 76 75 70 65
A U m5C G A G G G U m3C U C C A G A C Ψ A ΨA
60 59 57 56 55 50 47 43 40
A G C U m3C A G U C C U m22G C G G A A D D G mG D G A G (α4)C C G G U 5'
35 30 25 20 15 10


f
pA
pC
pU
pG
A 77
mA 67
Ψ 64
T 63
m3C 50
UmU 44,45
i6A 37
I 34
m3C 32
D 20
D 19
GmG 17,18
C 16
ac4C 12
g
D 20
G 9
A 68
mA 67
Ψ 64
T 63
m3C 50
UmU 44,45
C 34
m3C 32
D 20
D 19
GmG 17,18
D 16
ac4C 12
h
pY
pG
pA
A 77
mA 67
Ψ 64
T 63
m3C 50
UmU 44,45
i6A 37
I 34
m3C 32
D 20
D 19
GmG 17,18
C 16
ac4C 12
i
D 20
G 9
A 68
mA 67
Ψ 64
T 63
m3C 50
UmU 44,45
C 34
m3C 32
D 20
D 19
GmG 17,18
D 16
ac4C 12


j
pA
pC
pG
pI
pU
1
2

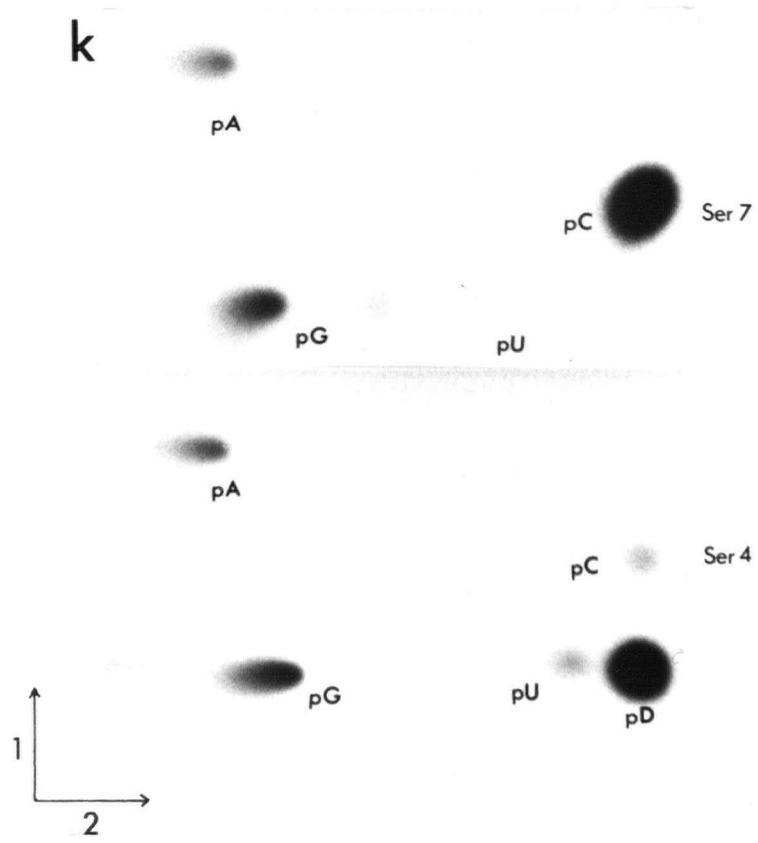

k
pA
pC
Ser 7
pG
pU
pA
pC
Ser 4
pG
pU
pD
1
2

Figure 6 - Gel "read-off" analysis of $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$.

(a): Samples of $tRNA_4^{Ser}$ (4.5 µg) and $tRNA_7^{Ser}$ (5 µg) were de-phosphorylated as described in Chapter II, then 5'-end labelled with polynucleotide kinase and [γ-^{32}P]ATP (30 µM; 450 Ci/mmole). Radiolabelled RNAs were separated by electrophoresis on a 20% polyacrylamide gel (4 hr. at 1200V), and located by autoradiography (1.5 min. using "no-screen" X-ray film).

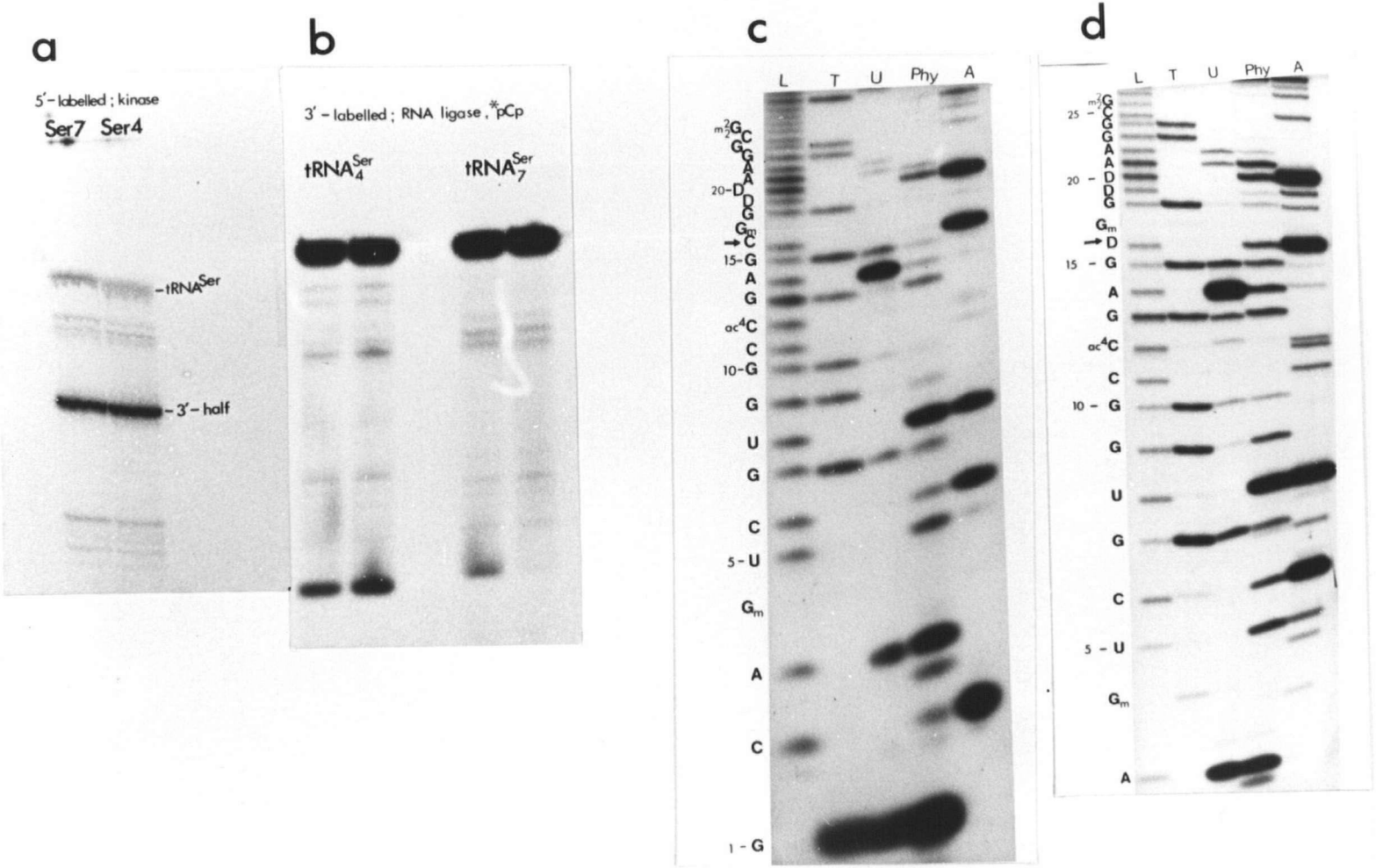
(b): Samples of $tRNA_4^{Ser}$ (0.9 µg) and $tRNA_7^{Ser}$ (1.0 µg) were labelled with RNA ligase and *pCp (1.2 µM; 320 Ci/mmole) as described in Chapter II. Radiolabelled RNAs were separated by electrophoresis on a 12% polyacrylamide gel (1.5 mm thick instead of 0.5 mm) for 15 hr. at 300V, and located by autoradiography (30 min. using "no-screen" X-ray film).

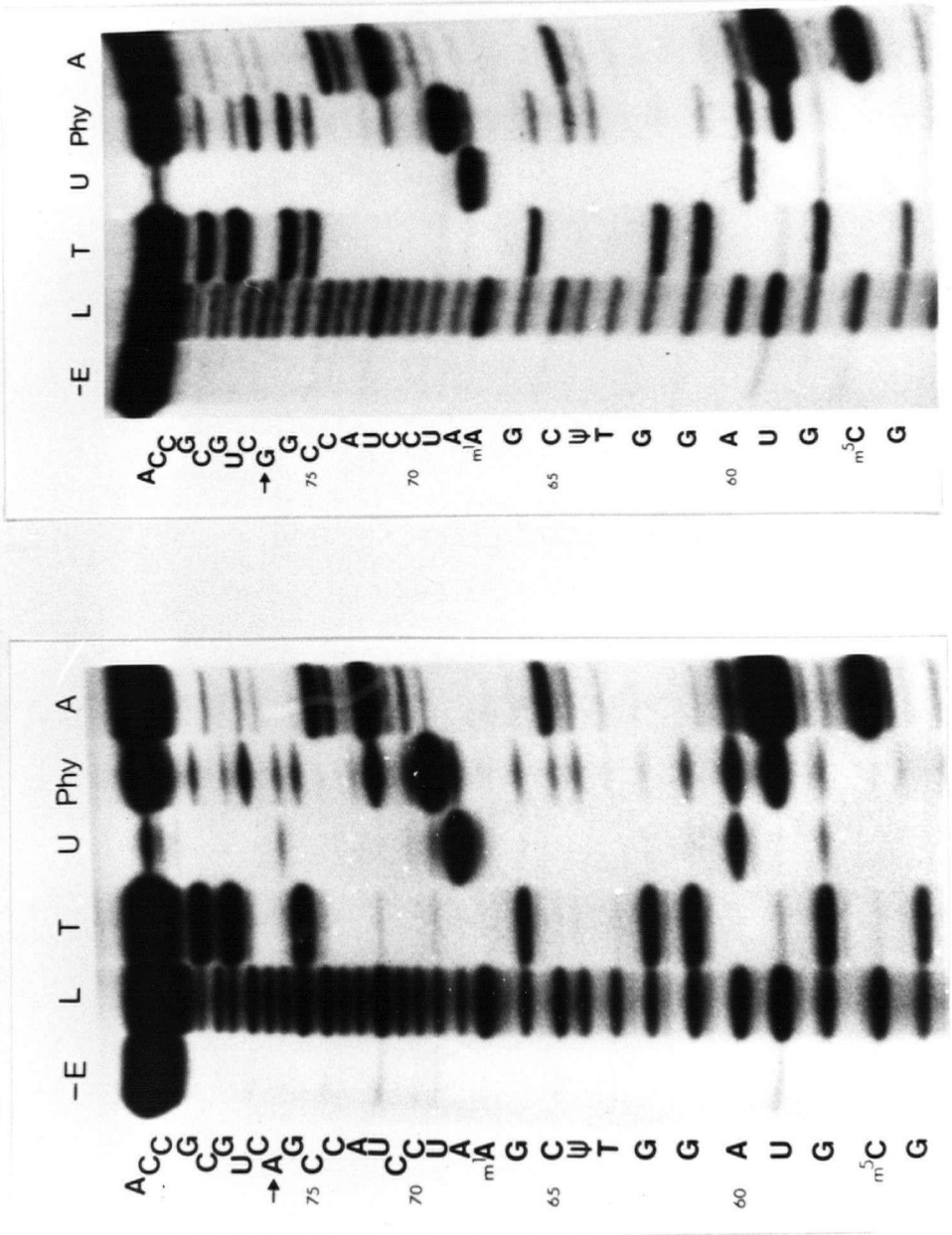
(c) and
(d): Partial enzymatic hydrolyses of [^{32}P]$tRNAs^{Ser}$ were performed as described in Chapter II, using the following amounts of ribonuclease activity per microgram of carrier tRNA: no enzyme (-E), 0.0067 unit RNase T_1 (T)/µg, 0.0067 unit RNase U_2 (U)/µg, 0.17 unit RNase Phy I (Phy)/µg, or 0.00067 unit RNase A (A)/µg. Reactions were for 15 min. at 55° (-E, T, A), 15 min. at 0° (U), 15 min. at room temperature (Phy), or 40 min. at 100° for the reference ladder (L). Twice as much [^{32}P]RNA was used for the L, Phy, and A slots as for the -E, T, and U slots. The T slots contained 7,000 Cerenkov cpm (c) and 11,000 Cerenkov cpm (d). Electrophoresis was for 2 hr. at 1500V on a 20% polyacrylamide gel (c) or for 2.8 hr. at 1800V on a 25% polyacrylamide gel (d). Autoradiography was for 4 days (c) or for 9 days (d) at -20° using Kodak XR-1 X-ray film.

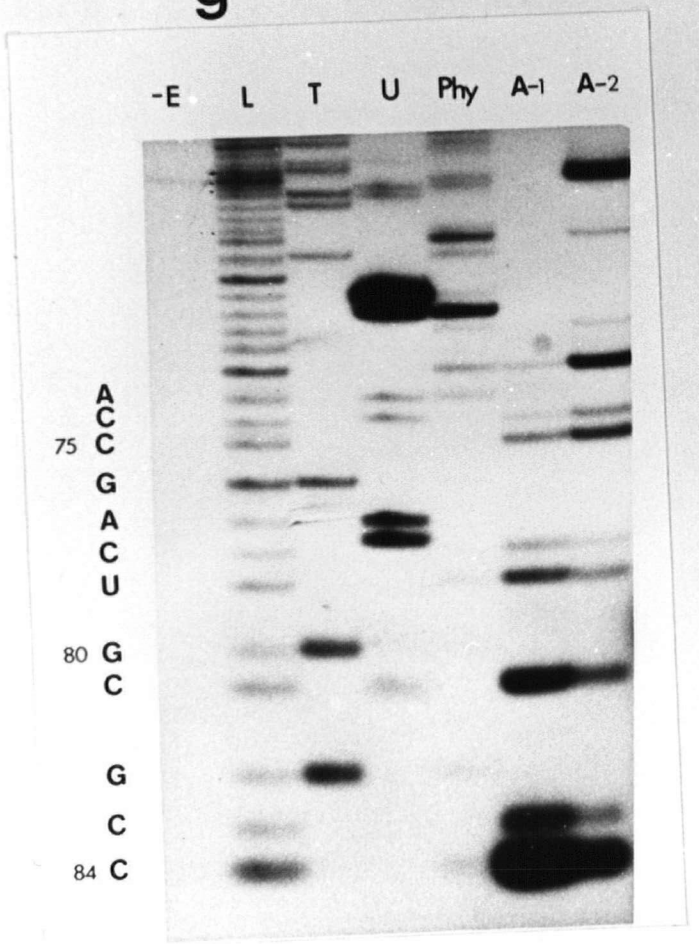
(e) and
(f): Partial hydrolyses were performed as described above for (c) and (d). The substrates were 5'-labelled 3'-half molecules, shown in part (a). The T slots contained 17,500 Cerenkov cpm of the 3'-half of $tRNA_7^{Ser}$ (e) or 7,000 Cerenkov cpm of the 3'-half of $tRNA_4^{Ser}$ (f). As before,

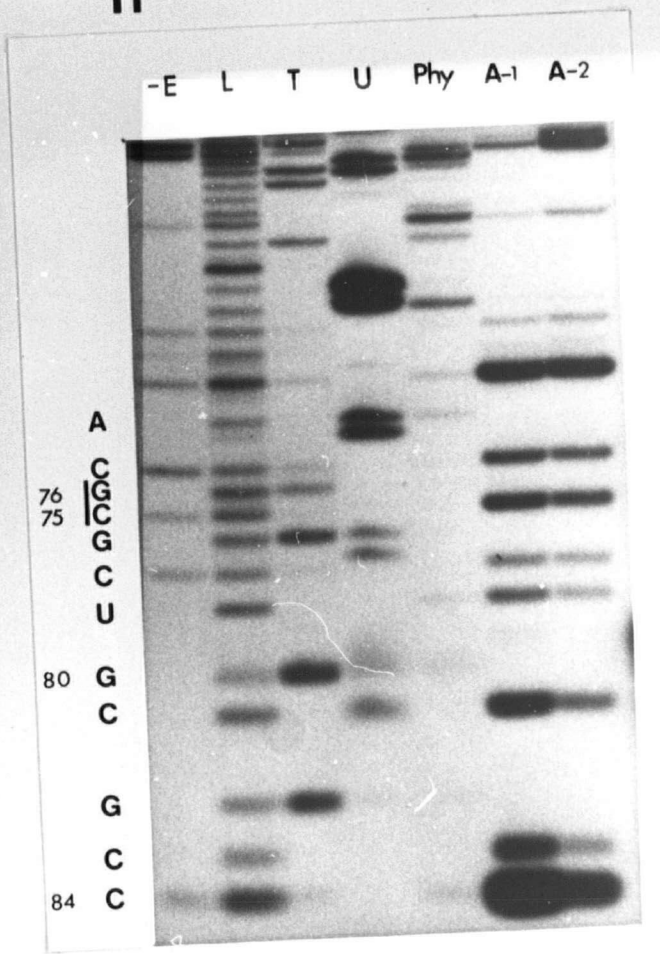
twice as much [^{32}P]RNA was present in the L, Phy, and A slots as in the -E, T, and U slots. Electrophoresis on 25° polyacrylamide gels was for 5 hr. at 1600V. Autoradiography was for 4 days (e) or 7 days (f) using Kodak "no-screen" X-ray film.

(g) and (h): Partial hydrolyses were performed using 2 x 10^{-3} unit RNase T_1/µg carrier tRNA (T), 2 x 10^{-3} unit RNase U_2/µg (U), 0.03 unit RNase Phy I/µg (Phy), 1 x 10^{-3} and 1 x 10^{-4} unit RNase A/µg (A-1 and A-2, respectively). Each slot contained 23,500 Cerenkov cpm (g) or 19,000 cpm/slot (h) except the reference ladders (L) which contained twice as much radioactivity. Electrophoresis was for 2.75 hr. at 1000V on 20% polyacrylamide gels. Autoradiography was for 37 hr. using Kodak "no-screen" X-ray film.


a
5'-labelled ; kinase
Ser7 Ser4
-tRNA^Ser
-3'-half
b
3' - labelled ; RNA ligase , *pCp
tRNA4^Ser
tRNA7^Ser
c
L T U Phy A
m2G
C
G
A
A
20-D
D
G
Gm
C
15-G
A
G
ac4C
C
10-G
G
U
G
C
5-U
Gm
A
C
1-G
d
L T U Phy A
m2G
C
G
A
A
20 - D
D
G
Gm
D
15 - G
A
G
ac4C
C
10 - G
G
U
G
C
5 - U
Gm
A


e
f
-E L T U Phy A
-E L T U Phy A


g
-E
L
T
U
Phy
A-1
A-2
A
C
C
75 C
G
A
C
U
80 G
C
G
C
84 C


h
-E
L
T
U
Phy
A-1
A-2
A
C
76 G
75 C
G
C
U
80 G
C
G
C
84 C

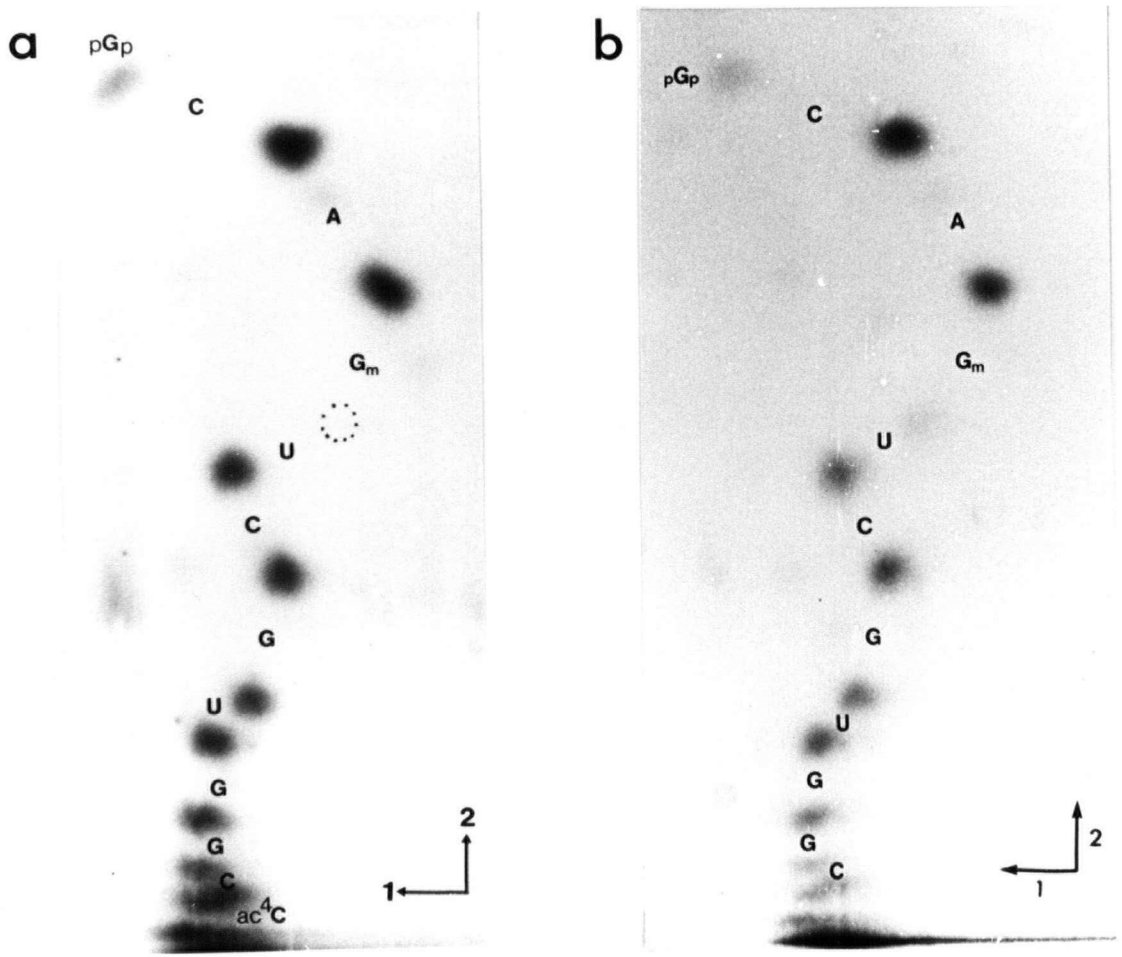
were identified by gel "read-off" analysis of $tRNA_7^{Ser}$ labelled at the 3'-terminus with *pCp by RNA ligase.

The statements above are fairly general in the sense that sequence analyses of $tRNA_{2b}^{Ser}$, $tRNA_4^{Ser}$, and $tRNA_5^{Lys}$ were carried out in much the same way. At this point, it is relevant to give a detailed description of the arguments used to arrive at the completed $tRNA_7^{Ser}$ sequence. My presentation will deal with points in the sequence where difficulty was encountered, moving generally from the 5'-end to the 3'-end. As the same arguments apply in large part to both $tRNA_4^{Ser}$ and $tRNA_{2b}^{Ser}$ analyses, I will discuss them much more briefly, emphasizing differences from $tRNA_7^{Ser}$.

The 5'-terminus of $tRNA_7^{Ser}$ is pG. White et al. found pGp in RNase T_2 digests of this tRNA (102). Gel "read-off" analysis of [5'-^{32}P] $tRNA_7^{Ser}$ (Fig. 6c) demonstrated the presence of G in the terminal position, as did chromatographic analysis of the *pN generated from [5'-^{32}P]tRNA by nuclease P_1 digestion (result not shown). Nucleotide analysis indicated two pG_m residues per $tRNA_7^{Ser}$ molecule (Table 1). One occurs at position 17 as shown below; the other was deduced to occur in position 4 based on the following considerations. The presence of a 2'-O-methylnucleoside at position 4 was shown by a gap in the ladders of gel "read-off" experiments (Fig. 6c) which also indicated that position 5 contains uridine. The mobility shift seen in "wandering spot" analysis of [5'-^{32}P]$tRNA_7^{Ser}$ (Fig. 7a) is consistent with the sequence G_mU or U_mG; however, the latter possibility was discarded based on the results of gel "read-off" experiments cited above. Therefore the sequence at positions 4 and 5 is G_mU. Position 12 contains ac^4C. In nucleotide analysis experiments (Fig. 4a), *pac^4C

Figure 7 - "Wandering spot" analysis of the 5'-termini of $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$.

[5'-^{32}P]$tRNA_7^{Ser}$ (a) and [5'-^{32}P]$tRNA_4^{Ser}$ (b) were hydrolyzed in 70% formamide for 60 min. at 100°. Samples (2 µl) contained 12,700 Cerenkov cpm (a) and 7900 Cerenkov cpm (b). Electrophoresis and homochromatography were performed as described in Chapter II. Autoradiography was for 16 days using "no-screen" X-ray film.


a
pGp
C
A
G_m
U
C
G
U
G
G
C
ac^4C
2
1
b
pGp
C
A
G_m
U
C
G
U
G
G
C
2
1

was identified by comparing its chromatographic properties with those of the commercially obtained nucleotide standard in solvents A and B (Chapter II, section A.8.g.). Analysis of *pNp's generated by hydrolysis of [^{32}P] oligonucleotides in 0.2M NaOH in a Stanley/Vassilenko experiment showed only C in position 12. This result would be expected with an alkali-labile nucleoside such as ac^{4}C (4, 157). A *pN with the correct mobility in solvent A was present in small quantities at this position in the same Stanley/Vassilenko experiment (Fig. 5e of ref. 43). In Stanley/Vassilenko analysis of $tRNA_{2b}^{Ser}$, the *pN in this position was clearly shown to be *pac^{4}C, based on comigration with the commercially obtained nucleoside 5'-phosphate in two solvent systems (Fig. 10c). The identity of the nucleosides present in position 12 of $tRNA_7^{Ser}$, $tRNA_4^{Ser}$, and $tRNA_{2b}^{Ser}$ is supported by the doublet pattern seen in the RNase A slot of gel "read-off" experiments for the three tRNAs. (This is seen most clearly for $tRNA_4^{Ser}$ in Fig. 6d, but compare with the equivalent position in $tRNA_7^{Ser}$, Fig. 6c, and $tRNA_{2b}^{Ser}$, Fig. 9a.) The doublets are not due to partial modification: only one band is present in the reference ladder. The doublets are not due to partial de-acetylation in the RNase A reaction mix, as larger fragments containing position 12 would show the same pattern. The doublets could be the result of efficient ring-opening of the 2', 3'-cyclic phosphate produced by initial RNase A-catalyzed cleavage. A similar doublet has been observed at a m^{5}C residue in yeast $tRNA_1^{Gly}$ (Fig. 8 of ref. 43). As the substituent in m^{5}C is not labile, increased rate of cyclic phosphate ring-opening by RNase A at modified cytosine residues is the favored explanation. Thus, the presence of doublet bands in the RNase A slot can be diagnostic for the presence of some modified cytosines.

Position 16 of $tRNA_7^{Ser}$ contains cytosine. This is clearly shown in Fig. 5k, where the *pN obtained from this position by nuclease P_1 digestion in a Stanley/Vassilenko experiment comigrated with pC on a two dimensional chromatograph. (The presence of cytidine here is discussed further below.) Gel "read-off" experiments show a gap in the reference ladder at position 17 (Fig. 6c), as does the ladder in a Stanley/Vassilenko experiment (Fig. 5a). Gel "read-off" experiments clearly show that G is present in position 18 (Fig. 6c). The G_m in position 17 was identified as $*pG_m$ by its chromatographic mobilities in solvents A and B on cellulose thin layer plates after extensive nuclease P_1 digestion (Methods). Thus, the slowly migrating species (Fig. 5d) derived by alkaline hydrolysis of a [5'-^{32}P] oligonucleotide is the dinucleoside triphosphate, $*pG_mGp$. Dihydrouridine residues in positions 19 and 20 were identified as *pN's by cellulose thin layer chromatography in solvents A and B (Fig. 5f and 5h).

Position 26 contains m_2^2G. The nucleotide there is apparently resistant to ribonuclease T_1 cleavage (Fig. 6c; ref. 158). The *pNp co-migrates with pCp (Fig. 5d). The nucleotide from position 26 of $tRNA_{2b}^{Ser}$, identified as pm_2^2G on the basis of chromatographic mobilities (Fig. 4c, 10c), is resistant to RNase T_1 cleavage and co-migrates with pCp on PEI cellulose developed with 0.80M ammonium sulfate. Therefore it was concluded that position 26 of $tRNA_7^{Ser}$ contains m_2^2G.

Identification of m^3C in position 32 in the anticodon loop, and in position 50 in the extra arm, could not be made on the basis of chromatographic mobilities as the *pN, since there were no relevant published values for this nucleotide (42). Also, no value was reported by Gupta

and Randerath (53) for the mobility of pm^3Cp on PEI cellulose in an ammonium sulfate solvent system. However, a number of observations indicate that the nucleoside at positions 32 and 50 is indeed m^3C. First, White et al. (102) identified a minor component of five serine tRNAs as m^3C', using the periodate oxidation and borotritiide-labelling method of Randerath et al. (159). In the solvent systems used for chromatography, m^3C' has distinctive mobilities, making its identification reasonably certain. Second, the pNp has a high mobility on PEI cellulose (Fig. 5d), similar to pm^1Ap and pm^7Gp which both contain positively charged bases. 3-methylcytidine carries a +1 charge under the conditions of chromatography (pK_a = 8.7). Third, the pN migrates similarly to pm^1A, pm^7G, and pC on two dimensional cellulose thin layer plates (Fig. 4a). Fourth, this nucleoside is resistant to ribonuclease A cleavage (Fig. 6c). RNase A does not cleave at cytosine residues modified at N-3 of the pyrimidine ring (160). Fifth, the "jump" between bands in the reference ladder of "read-off" experiments due to addition of this nucleotide is large, like that for a purine nucleotide or one containing a positively charged base. For these reasons, the nucleoside at positions 32 and 50 was identified as m^3C.

Inosine in position 34 (the "wobble" position) was identified as *pI by co-chromatography in solvents A and B with commercially obtained inosine 5'-phosphate (Fig. 5j). The chromatographic mobilities determined for $*pi^6A$ (position 36), obtained in a Stanley/Vassilenko experiment, in solvents A and B were the same as for commercially obtained pi^6A standard (Fig. 5f,5h). Pseudouridine in positions 39 and 64 chromatographed as expected for pψ in solvents A and B (Fig. 5f,5h), and as expected for pψp on PEI cellulose plates in the ammonium sulfate solvent (Fig. 5d; refs.

43,53). It was cleaved like a U by ribonucleases PhyI and A in gel "read-off" experiments (Fig. 6e).

The sequence at positions 44-48 was deduced to be U_mUCCC. The presence of a 2'-O-methylnucleoside at position 44 was shown by a gap in the reference ladder of gel "read-off" experiments. The nucleoside 5'-phosphate released by extensive nuclease P_1 digestion in a Stanley/Vassilenko experiment has the same chromatographic mobilities in solvents A and B as standard pU_m. Nucleotide analysis indicated the presence of pU_m at less than one molecule per $tRNA_7^{Ser}$ (Fig. 4a, Table 1), and both pG_m residues (Table 1) were accounted for at positions 4 and 17. Thus position 44 contains U_m. Gel "read-off" experiments indicate the nucleoside at position 45 is U. Consistent with this, the chromatographic mobility of the dinucleoside triphosphate on PEI cellulose in 0.80M ammonium sulfate is very nearly the same as that determined for pU_mpUp in a similar chromatographic system by Gupta and Randerath (53). C46, C47, and to a lesser extent C48 could be identified in gel "read-off" experiments, while C48, C47, and sometimes C46 could be obtained as separate bands on polyacrylamide gels in Stanley/Vassilenko experiments (Fig. 5a), though bands in this region tended to compress during electrophoresis, presumably due to persistent secondary structure (135). (Such compression was particularly troublesome during the earlier sequencing experiments in the work reported here, when we did not have a power supply that would maintain a sufficiently high amperage to run polyacrylamide gels at elevated temperatures using standard buffers.) The possibility remains that position 45 contains ψ rather than U. Nucleotide analysis indicated 3.2 $p\psi$ per tRNA molecule estimated relative to pT (Table 1). Based on the methods used here that possibility cannot be excluded. However, quantitation of

modified nucleotides in a tRNA is subject to rather large errors at times due to variations in polynucleotide kinase labelling efficiency. Lacking more direct evidence for the presence of ψ at this position, the sequence at positions 44-48 was deduced to be U_mUCCC.

Nucleotides 56-58 were poorly labelled by polynucleotide kinase in Stanley/Vassilenko experiments (Fig. 5a), but could be tentatively read as Gm^5CG. Nucleotide analyses of both $tRNA_7^{Ser}$ and $tRNA_{2b}^{Ser}$ indicated the presence of pm^5C, which migrates very near to a pm^7G standard in both solvents A and B on cellulose thin layer plates. In $tRNA_{2b}^{Ser}$, this pm^5C was located in position 57 (Fig. 10c). The pm^5Cp from the same position of $tRNA_{2b}^{Ser}$ migrated slightly ahead of pCp on PEI cellulose plates in the ammonium sulfate system, as did the putative pm^5Cp from $tRNA_7^{Ser}$. Supporting this interpretation, gel "read-off" experiments yielded clear patterns read as GCG (Fig. 6e). In one case a doublet pattern like that discussed above for ac^4C was seen (results not shown). "Wandering spot" analysis of a [^{32}P]oligonucleotide obtained in a Stanley/Vassilenko experiment indicated the presence of the sequence AGCGUAGG here. Taken together, these data are all consistent with the sequence Gm^5CG at positions 56-58.

Ribothymidine in position 63 and 1-methyladenosine in position 67 were both identified as *pN's by their chromatographic mobilities on cellulose thin layer plates in solvents A and B (Fig. 5f,5h). As pNp's, *pTp comigrates with pUp (Fig. 5d), while $*pm^6Ap$ (derived from $*pm^1ApNp \cdots CCA$ during alkaline hydrolysis; Chapter II, A.8.e.) migrates somewhat faster than pAp (Fig. 5d). Both T and m^1A are resistant to ribonuclease cleavage in "read-off" experiments. A large "jump" in the reference ladder

is seen on addition of m^1A to the 5' end in such experiments. The distinctive pattern found at the sequence $GT\psi CGm^1AA$ common in tRNAs is shown in Fig. 6e.

The sequence from positions 68-82 was unambiguously determined in Stanley/Vassilenko and gel "read-off" experiments. The 3'-terminal CCA sequence (positions 83-85) was determined by gel "read-off" analysis of $tRNA_7^{Ser}$ labelled at the 3'-end with *pCp by RNA ligase. C83 and C84 could be read directly from autoradiographs (see Fig. 6g). A85 was identified by "nearest neighbor" analysis of *pCp-labelled $tRNA_7^{Ser}$, as follows. Radiolabelled $tRNA_7^{Ser}$ was digested to nucleoside 3'-phosphates by RNase T_2 (resulting in transfer of the [5'-^{32}P] phosphoryl group of *pCp to its neighbor on the 5' side) the nucleotides separated by cellulose thin layer chromatography in solvent A (together with Np standards), and the [^{32}P] phosphoryl group located in Ap* by autoradiography (results not shown). It should be noted that the presence of a 3'-terminal adenosine can also be deduced directly from gel "read-off" experiments, where autoradiographs show doublet band patterns at A residues on digestion of the [3'-^{32}P] tRNA with RNase U_2 (see Fig. 6g). This pattern, which can be misleading as the lower band (sometimes the stronger) occasionally comigrates with the next smaller reference ladder band, results from a high percentage of doubly cleaved molecules containing 3'-terminal CCAp* rather than CCA*pCp. Thus, while "nearest neighbor" analysis should be performed routinely as part of the sequencing protocol, it is not really necessary in every case encountered.

The completed $tRNA_7^{Ser}$ sequence, determined as described above, is presented as a "cloverleaf" structure in Fig. 8.

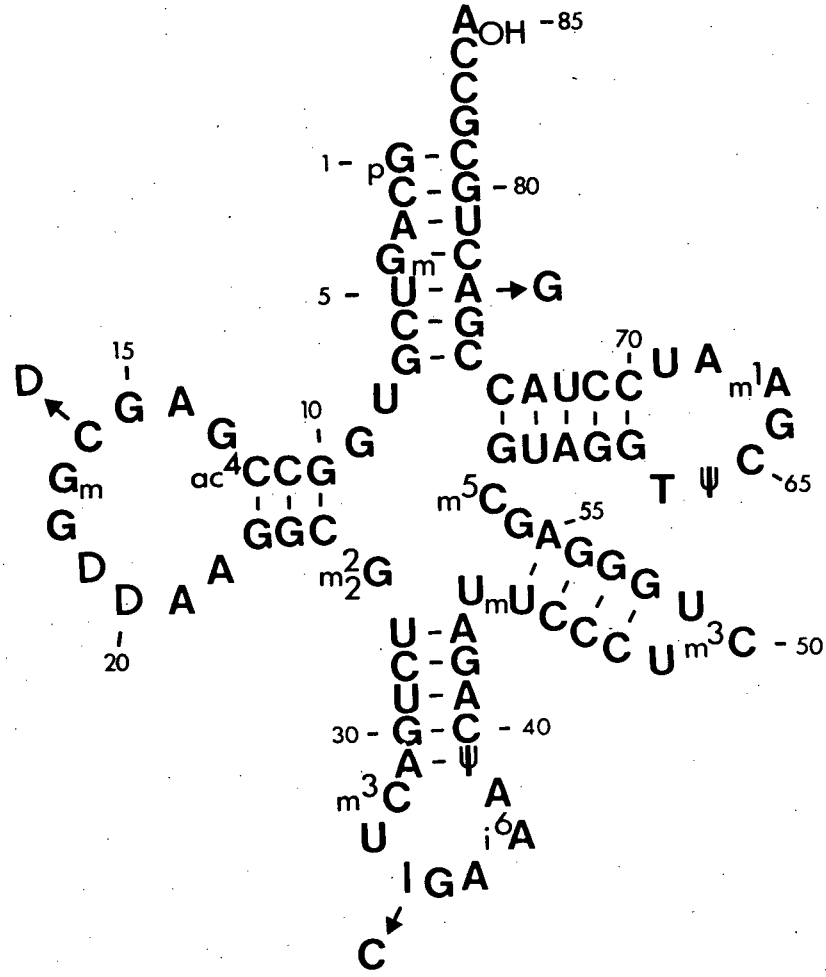


Figure 8 - Cloverleaf structure of $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$.

It should be noted here that the criteria applied to deduce the tRNA sequences presented in this chapter are quite rigorous, and sequences obtained by this general strategy are apparently accurate, as DNA sequences 100% homologous to $tRNA_7^{Ser}$ (see Chapter IV) and $tRNA_5^{Lys}$ (Chapter V) have been found. This point is an important one in the interpretation of DNA sequence analyses presented in Chapter IV.

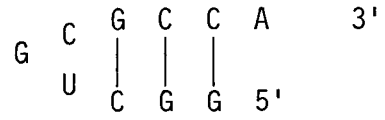
B.2. Sequence analysis of $tRNA_4^{Ser}$

The nucleotide sequence of $tRNA_4^{Ser}$ was determined in an almost identical fashion to that of $tRNA_7^{Ser}$. Generally, the same experiments were carried out in parallel with the two tRNAs. Nucleotide analysis is presented in Fig. 4b; quantitation of nucleotides is in Table I. The results of Stanley/Vassilenko analysis are shown in Figures 5b, 5c, 5e, 5g, 5i. The results of representative gel "read-off" analyses of [5'-^{32}P] $tRNA_4^{Ser}$, a 5'-labelled half-molecule containing nucleotides 36-85, and [3'-^{32}P]$tRNA_4^{Ser}$ labelled by RNA ligase are presented in Figure 6d, 6f, and 6h. The 5'-terminal sequence identified in a "wandering spot" experiment is shown in Fig. 7b. The entire sequence was covered two or more times. All nucleotides except the 3'-terminal CCA were determined by two or more independent methods.

The nucleotide sequence of $tRNA_4^{Ser}$ is very similar to that of $tRNA_7^{Ser}$, differing at only three out of 85 positions. The similarity between the two sequences is readily seen by comparing equivalent gel "read-off" experiments. For example, patterns obtained from 5'-terminal sequences (Figures 6c, 6d) or from 3'-terminal sequences (Figures 6e, 6f)

are nearly identical. However, each is distinct, differing at positions indicated by arrows in the figures. The three positions differing between $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ are 16, 34, and 77. Position 16 contains a D in $tRNA_4^{Ser}$ but C in $tRNA_7^{Ser}$, as is clearly seen in Fig. 5k. Position 34 is the "wobble" position of the anticodon. While $tRNA_7^{Ser}$ contains inosine there (Fig. 5j), the same position in $tRNA_4^{Ser}$ contains cytidine (Figures 5e, 5g, 5i).

The greatest difficulties encountered in analysis of the $tRNA_4^{Ser}$ sequence were in deducing the sequence at and around nucleotide 77. The sequence is easily read as $A_{74}\underline{CCGG}CUG_{80}$ in the autoradiograph of a "read-off" gel shown in Fig. 6f. However, in gel "read-off" analysis of [3'-^{32}P] $tRNA_4^{Ser}$, the sequence was read as $A_{74}\underline{CGCG}CUG_{80}$ (Fig. 6h). In a Stanley/Vassilenko experiment, the sequence was read as $AC\underline{C-G}CUG$ (Fig. 5b). One band is missing from the ladder relative to $tRNA_7^{Ser}$ (compare to Fig. 5a). In a separate Stanley/Vassilenko experiment, a doublet was seen at this position (Fig. 5c). The 5'terminal nucleotide of each band of the doublet was *pGp (results not shown). These discrepancies can be resolved in the following way. The gel "read-off" experiment shown in Figure 6f is very clear and easily read. The band pattern for $tRNA_4^{Ser}$ here matches that for $tRNA_7^{Ser}$ (Fig. 6e) at all positions except 77, where the two tRNAs contain G and A, respectively. All problems encountered at this region of the $tRNA_4^{Ser}$ sequence involved the electrophoretic behavior of short [^{32}P] oligonucleotides containing the 3'-terminal GCCA sequence. The sequence $\underline{GCC}A$ is complementary to GGC at positions 76-78. Thus, the 3'-terminal sequence containing nucleotides 76-85 is capable of forming a hairpin structure containing three G-C base pairs and a three nucleotide loop:



Under non-denaturing electrophoretic conditions oligonucleotides containing this hairpin might migrate through polyacrylamide gels at an anomalously high rate, resulting in a pattern that does not represent the true sequence. As noted earlier, the power supply we were using at this time did not maintain sufficiently high amperage to run polyacrylamide gels at elevated temperatures (indeed, the problem cited earlier involved the extra arm of $tRNA_7^{Ser}$, a hairpin structure containing three consecutive G-C base pairs and a three nucleotide loop.) However, such considerations would not be relevant in gel "read-off" analysis of [5'-^{32}P] $tRNA_4^{Ser}$ (or the 5'-end labelled 3'-half molecule, as shown in Fig. 6f), since the 3'-terminal GCCA involved in forming such a hairpin would not be present in radiolabelled oligonucleotides ending at G77.

The completed nucleotide sequence of $tRNA_4^{Ser}$ is presented as a "cloverleaf" structure in Fig. 8. The arrows point to the nucleotide changes in $tRNA_4^{Ser}$ distinguishing it from $tRNA_7^{Ser}$.

B.3. Sequence analysis of $tRNA_{2b}^{Ser}$

The nucleotide sequence of $tRNA_{2b}^{Ser}$ was determined by the same methods used to analyze $tRNA_7^{Ser}$ and $tRNA_4^{Ser}$, and also by a variation of the Stanley/Vassilenko method described by Tanaka et al. (ref. 54; Chapter II, A.9.). The data obtained were not so extensive as for $tRNA_{4,7}^{Ser}$ (above) but nonetheless allowed an unambiguous sequence to be assigned.

In contrast to $tRNAs^{Ser}_{4,7}$, nucleotide analysis clearly indicated the presence of pm^2_2G in $tRNA^{Ser}_{2b}$, as well as pt^6A and a nucleotide that is probably pmt^6A (Fig. 4c). Nucleotide content in $tRNA^{Ser}_{2b}$ was estimated; these values are found in Table I. The entire sequence was covered in gel "read-off" analyses of [5'-^{32}P] $tRNA^{Ser}_{2b}$, a 5'-labelled 3'-half molecule obtained from polynucleotide kinase labelling reactions, and [3'-^{32}P] $tRNA^{Ser}_{2b}$. Representative autoradiographs from such experiments are shown in Figures 9a, 9b, and 9c. The nucleotide sequence from positions 4-74 was determined in Stanley/Vassilenko experiments (Fig. 10). In those experiments, pac^4C12, pG_m17, and pU_m44 were identified by their co-migration with commercially obtained nucleoside 5'-phosphate standards on cellulose thin layer chromatography in solvents A and B; pm^2_2G26 and pm^5C57 were clearly identified on the basis of chromatographic mobilities in the same solvents (as mentioned in section B.1. above dealing with $tRNA^{Ser}_7$); and other modified nucleotides were identified by their mobilities as pN's (Fig. 10c), as described above. Position 37 adjacent to the anticodon contains a nucleotide with chromatographic mobilities expected for pt^6A (ref. 42) and an additional species that is likely to be pmt^6A (Fig. 4c and 10d). Since no mobility values for pmt^6A are available, this identification is indirect, based on homology with rat liver $tRNA^{Ser}_3$. The rat liver tRNA sequence is nearly identical to that of $tRNA^{Ser}_{2b}$ (98%), including modifications elsewhere in the sequence. This high degree of conservation is likely maintained at position 37 as well, where rat liver $tRNA^{Ser}_3$ contains mt^6A (7). White et al. (102) found an unknown nucleotide in $tRNA^{Ser}_2$ with properties similar to mt^6Ap, but did not see t^6Ap. Sequence

Figure 9 - Gel "read-off" analysis of $tRNA_{2b}^{Ser}$.

(a) [5'-^{32}P]$tRNA_{2b}^{Ser}$ was prepared by de-phosphorylating 2 µg of purified tRNA and labelling with polynucleotide kinase and [γ-^{32}P]ATP (40 µM, 1000 Ci/mmole). Partial digestions were performed as described for Figure 6c. Control (-E, not shown), RNase T_1 (T), and RNase U_2 (U) slots contained 11,000 Cerenkov cpm while RNase Phy I (Phy), RNase A (A) and reference ladder (L) slots contained twice that amount (22,000 cpm). Electrophoresis on 20% polyacrylamide gels was for 2 hr. at 1800V. Autoradiography was for 8 days at -20° using Kodak "no-screen" X-ray film.

(b) The 5'-labelled 3'-half molecule including nucleotides 35-85 of $tRNA_{2b}^{Ser}$ was digested as in part (a) above. Control (-E, not shown), RNase T_1 (T), and RNase U_2 (U) slots contained 6,500 Cerenkov cpm, while the RNase Phy I (Phy), RNase A (A) and reference ladder (L) slots contained twice that amount. Electrophoresis was at constant power (30 watts) until the BB dye reached the bottom of the gel. Autoradiography was for 6 days at -20° using Kodak "no-screen" film.

(c) [3'-^{32}P]$tRNA_{2b}^{Ser}$ was prepared by labelling 1 µg $tRNA_{2b}^{Ser}$ with RNA ligase and 1.3 µM *pCp (~ 100 Ci/mmole). Partial hydrolyses were as described for part (a) above. -E (not shown), T, and U slots contained 4100 Cerenkov cmp, while the Phy, A, and L slots contained twice that amount. Electrophoresis was for 2.5 hr. at 1600V on a 25% polyacrylamide gel. Autoradiography was for 10 days at -20° using Kodak "no-screen" X-ray film.

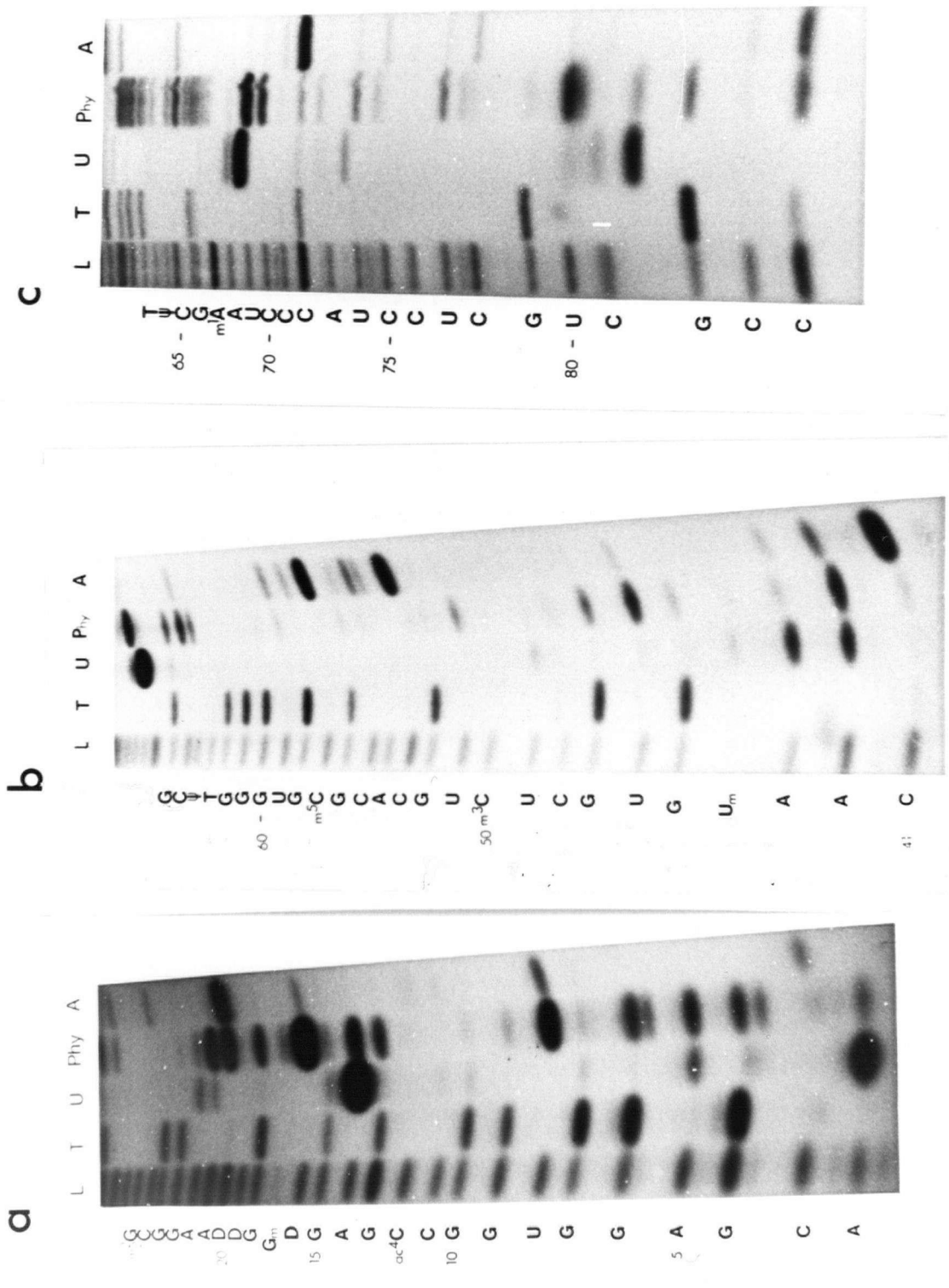

a
L T U Phy A
b
L T U Phy A
c
L T U Phy A

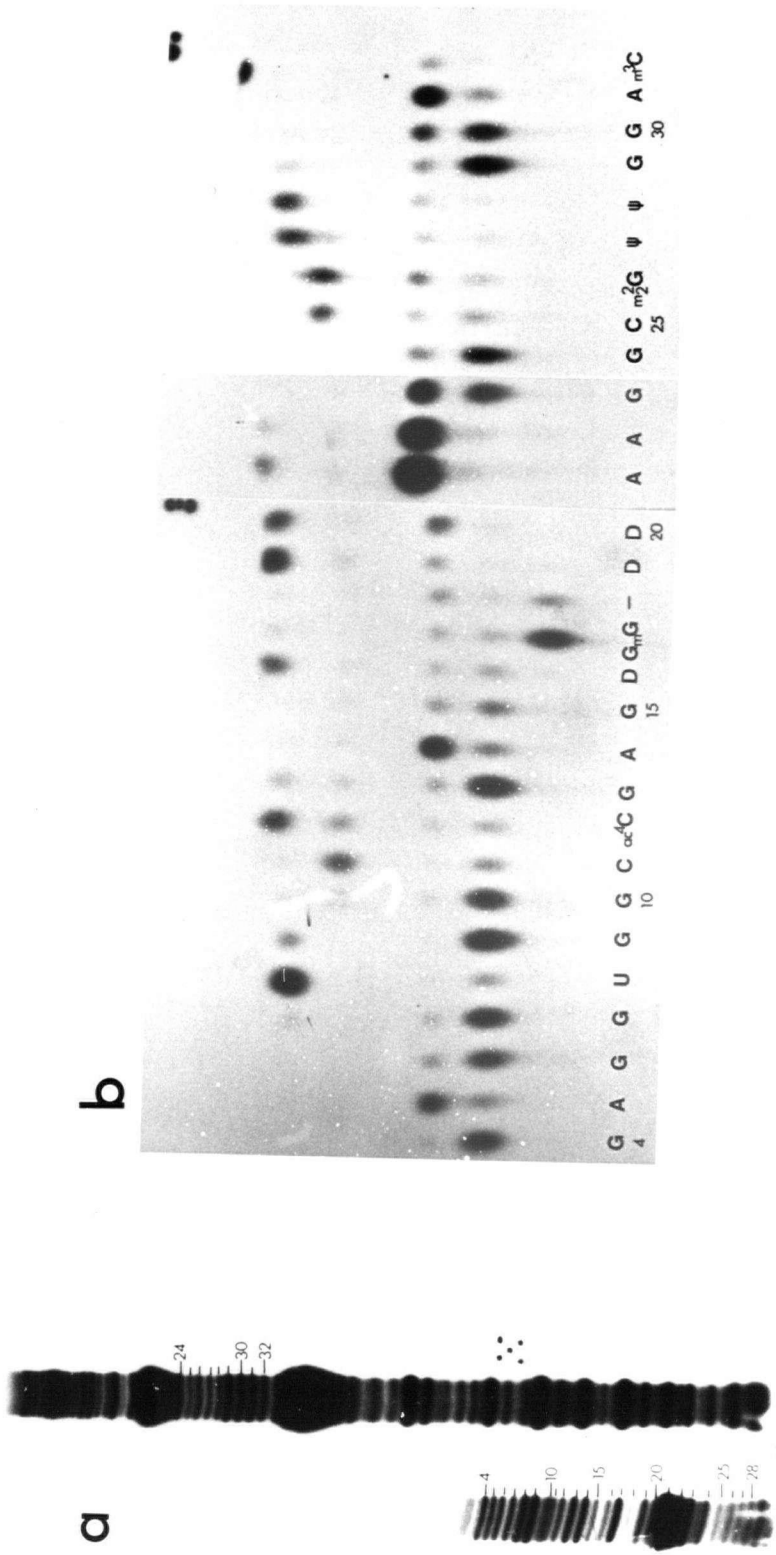
Figure 10 - Stanley/Vassilenko sequence analysis of $tRNA_{2b}^{Ser}$.

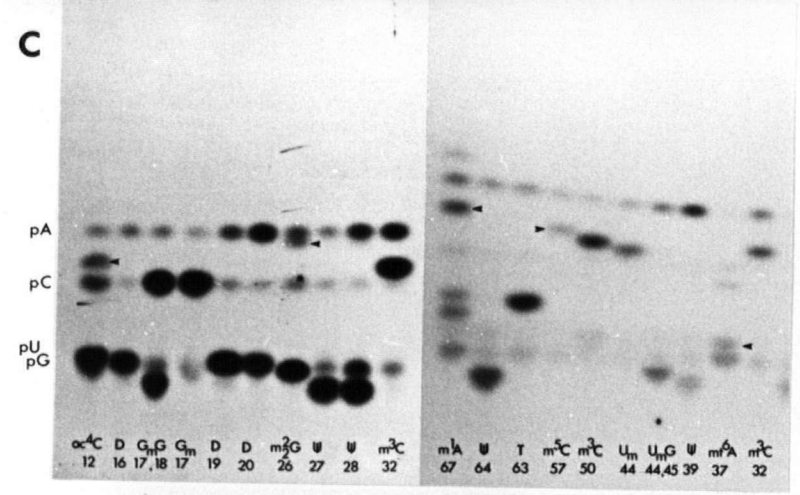
(a) Samples of purified $tRNA_{2b}^{Ser}$ (1 µg/slot) were partially hydrolyzed in 8 mM NH_4OAc (pH 4.6) in a sealed capillary tube, radiolabelled as described in Chapter II, then separated by electrophoresis on a 12% polyacrylamide gel as shown. Electrophoresis (at 20 watts, constant) was for 4.7 hr. or for 2.5 hr. Autoradiography was for 11 hr. at -20° using Kodak XR-1 X-ray film.

(b) The 5'-terminal nucleotides (*pNp's) of [^{32}P] oligonucleotides separated by polyacrylamide gel electrophoresis were analyzed by PEI cellulose chromatography as described in Chapter II. Nucleotides are represented by symbols for the bases, with positions in the tRNA sequence (Figure 12) indicated below.

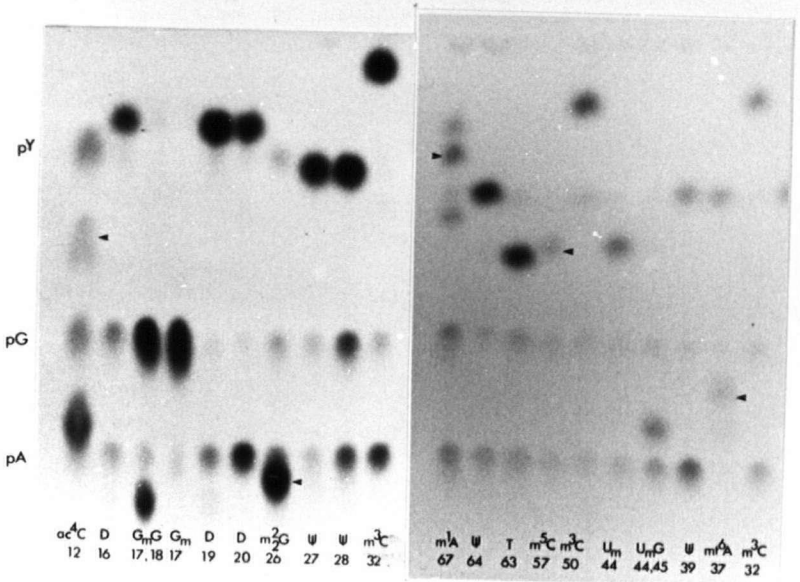
(c) The 5'-terminal nucleotides (*pN's) of some [^{32}P] oligonucleotides were analyzed by chromatography on cellulose thin layer plates developed either in solvent A or in solvent B as described in Chapter II. The nucleotides are represented by symbols for the bases, with positions in the tRNA sequence (Figure 12) indicated below.

(d) The *pN from position 37 adjacent to the $tRNA_{2b}^{Ser}$ anticodon was analyzed by two dimensional cellulose thin layer chromatography as described in Chapter II. The sample contained less than 200 Cerenkov cpm. Autoradiography was for 18 days at -20° using Kodak XR-1 X-ray film.

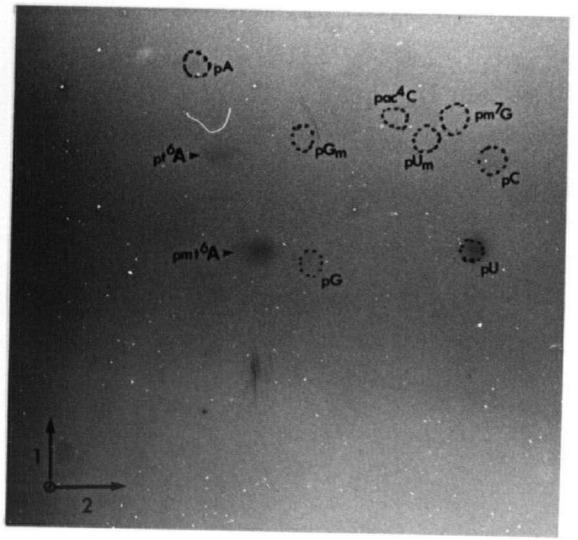

a
24
30
32
4
10
15
20
25
28
b
G A G G U G G C $_{ac}^{4}$C G A G D G$_m$G – D D A A G G C m$_2^2$G Ψ Ψ G G A $_m^3$C
4
10
15
20
25
30



Solvent A



Solvent B



analysis of $tRNA_{2b}^{Ser}$ by the similar method of Tanaka et al. (54) showed a single spot at this position, with an electrophoretic mobility of the pNp consistent with that expected for pt^6Ap or pmt^6Ap (Fig. 11b). Whether the heterogeneity observed at position 37 in Stanley/Vassilenko analysis reflects partial modification of t^6A to mt^6A in vivo or partial breakdown at some stage of sample preparation is not clear. However, I have assigned mt^6A to position 37, with the reservation that significant amounts of t^6A may be present there in $tRNA_{2b}^{Ser}$ in vivo.

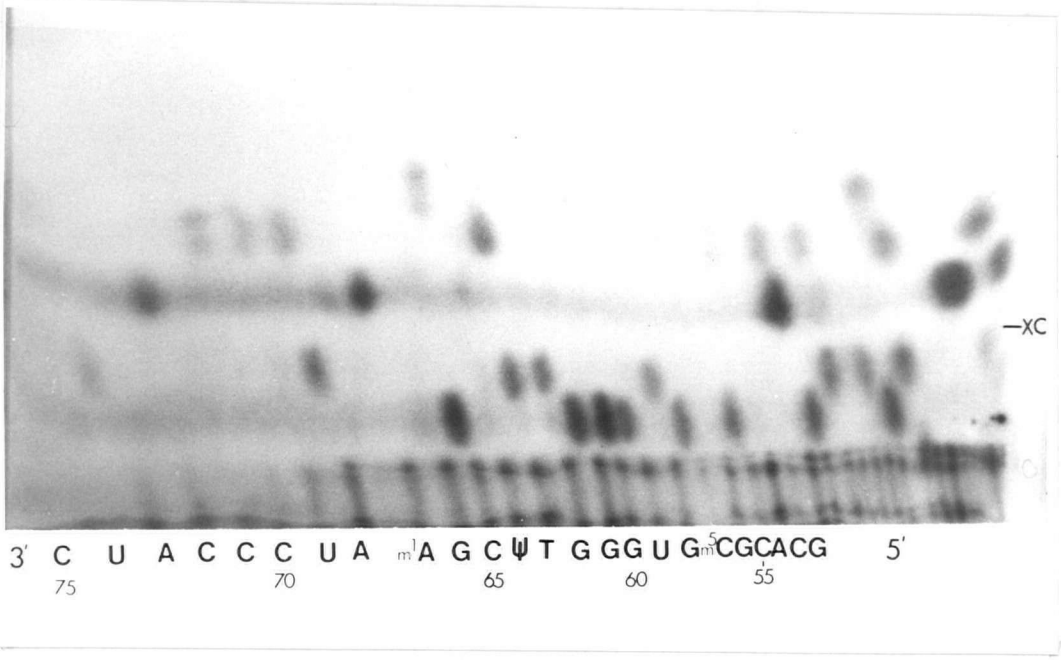
Use of ribonuclease T_2 in Stanley/Vassilenko experiments rather than 0.2M NaOH allowed the chromatographic properties of pDp and pm^1Ap (as well as pm^7Gp; Chapter V) in 0.8 M ammonium sulfate on PEI cellulose to be established. The mobilities of a number of pNp's from serine tRNAs and from $tRNA_5^{Lys}$ (Chapter V) are shown in Table 2.

The nucleotide sequence of $tRNA_{2b}^{Ser}$ from positions 13-75 was analyzed by the method of Tanaka, Dyer, and Brownlee (54) as well as by Stanley/Vassilenko and gel "read-off" methods. Experiments using the method of Tanaka et al. gave very clear results (Fig. 11) allowing most of the $tRNA_{2b}^{Ser}$ sequence to be read unambiguously. In these experiments, two aliquots of the same sample were electrophoresed in adjacent lanes of a polyacrylamide gel. One lane was used for these two-dimensional "read-off" experiments, while individual bands from the other lane were located by autoradiography, excised and treated as usual in my Stanley/Vassilenko protocol. Only the experiment allowing identification of modified nucleotides from the 3' half of $tRNA_{2b}^{Ser}$ as the *pNs is shown (Fig. 10c, 10d), as the sequence is otherwise more clearly read from the two-dimensional arrays in Fig. 11.

Figure 11 - Two-dimensional "read-off" sequence analysis of $tRNA^{Ser}_{2b}$; a variation on Stanley/Vassilenko.

Two samples of ~ 0.5 µg each gel-purified $tRNA^{Ser}_{2b}$ (hydrolyzed 10 min. at 100° in 98% formamide) were labelled with polynucleotide kinase and 4 µM [γ-^{32}P]ATP (2000 Ci/mmole). The [^{32}P] oligonucleotides were separated by electrophoresis on 20% polyacrylamide gels at a constant power of 18 watts until the XC dye had migrated to 45% (a) or 95% (b) the length of the gel. After transfer, digestion with RNase T_2 *in situ*, and electrophoresis in the second dimension (3 hr. at 250V), the DEAE-cellulose plates were air-dried and exposed to "screen" X-ray film for 16 hr.

a.


—XC
3' C U A C C C U A m1A G C Ψ T G GG U Gm5CGCACG 5'
75 70 65 60 55

b.

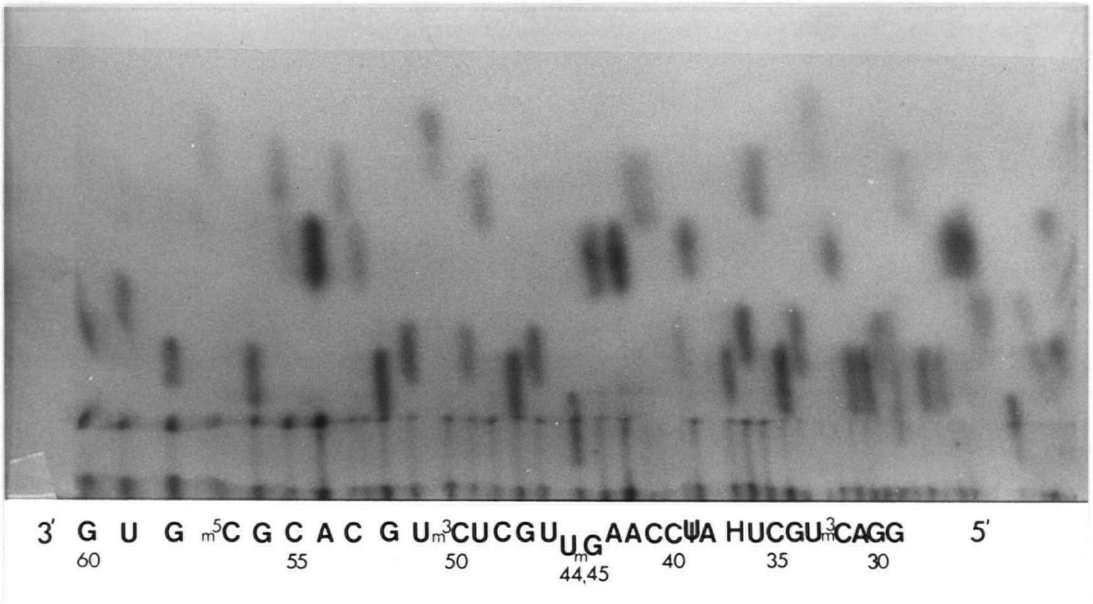

3' G U G m5C G C A C G U m3CUCGU m UG AACCΨA HUCGU m3CAGG 5'
60 55 50 44,45 40 35 30

Table 2 - Mobilities of nucleoside-5', 3'-bisphosphates on PEI cellulose[a]

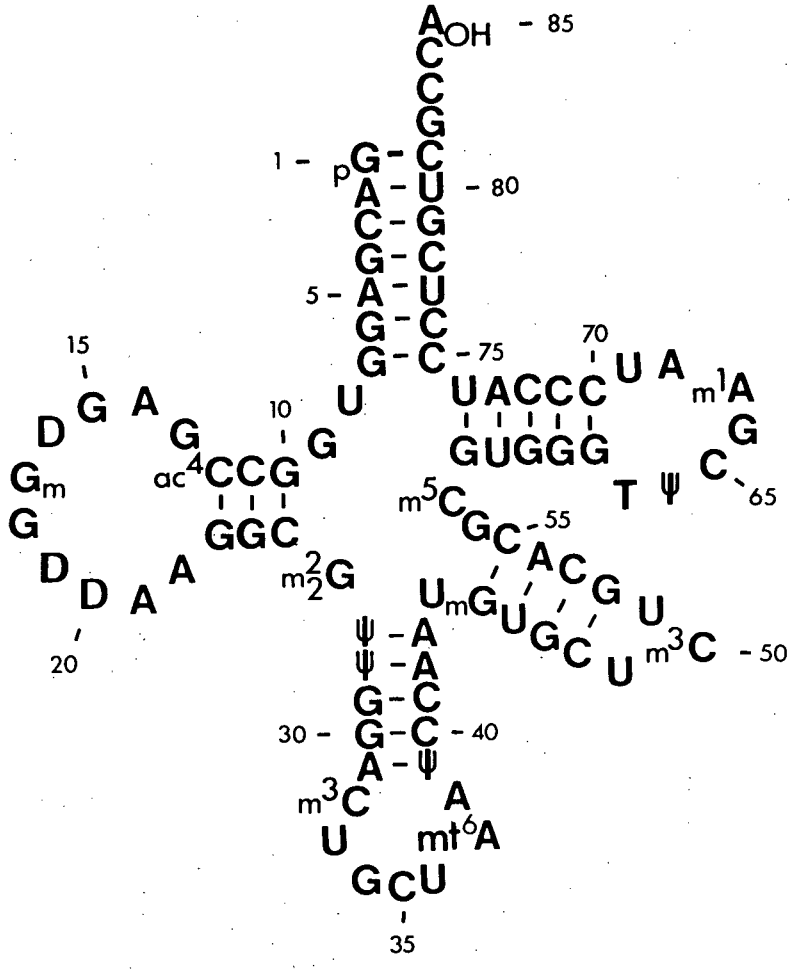
pNp	R_{pUp}[b]
pUp	1.00
pTp	1.00
pψp	0.94
pDp	1.06
pAp	0.66
pm^6Ap	0.75
pi^6Ap	0.52, 0.61[c]
pIp	0.84
pCp	0.83 = 3' (0.79)[d]
pm^3Cp	1.12
pGp	0.47 = 3' (0.40)[d]
pm^7Gp	1.08
pm_2^2Gp	0.83
pm^5Cp	0.85
pm^2Gp	0.61
pac^4Cp	1.04
pm^1Ap	1.12

a - Developed with 0.80 M ammonium sulfate (pH 5.3) at room temperature.

b - Mobility relative to pUp.

c - Two spots, for the 2' and 3'-isomers, are found.

d - Mobility value for the 2'-isomer.



$tRNA^{Ser}_{2b}$

Figure 12 - Cloverleaf structure of $tRNA^{Ser}_{2b}$.

The 3'-terminal CCA of $tRNA_{2b}^{Ser}$ was determined by gel "read-off" analysis of 3'-end labelled tRNA (Fig. 9c), while sequence slightly further 5' was determined by analysis of both 5'- and 3'-end labelled RNAs (Fig. 9b and 9c). Possible ambiguities in interpretation of "read-off" experiments between positions 81 and 73 were clarified by "wandering spot" analysis of 3'-end labelled $tRNA_{2b}^{Ser}$ (results not shown).

The complete nucleotide sequence of $tRNA_{2b}^{Ser}$ is presented as a "cloverleaf" structure in Fig. 12.

C. Discussion

C.1. Sequences of individual serine tRNAs from D. melanogaster

The nucleotide sequences of $tRNA_{2b}^{Ser}$, $tRNA_{4}^{Ser}$, and $tRNA_{7}^{Ser}$ from Drosophila melanogaster are unremarkable. Each fits well with the generalized "cloverleaf" structure of Rich and RajBhandary (Ref. 8; Fig. 2 in Chapter I above) with respect to stem lengths, invariant and semi-invariant nucleosides.

C.2. Recognition of Drosophila $tRNAs^{Ser}$ by seryl-tRNA synthetase

The features of Drosophila serine tRNAs recognized by seryl-tRNA synthetase must be present in all three tRNAs examined here, regardless of anticodon. There are many homologous positions in all three tRNA sequences which could be recognition sites for this enzyme. I suggest that the minimum number of recognition sites required can be estimated, and some potential recognition sites identified, in the following way. It

may be that synthetase recognition of cognate versus non-cognate tRNAs can be expressed as a set of yes-or-no decisions at specific tRNA structural "determinants". If so, each of the twenty isoacceptor tRNA families would necessarily contain at least five "determinants" (yes-or-no yields two alternatives at a "determinant site"; $2^5 = 32 > 20$). Each "determinant" would be independently recognized by synthetases as correct or discarded as incorrect. Such "determinants" might comprise the presence of a particular base (standard or modified) in a particular position in the three dimensional configuration; or more subtle features, such as the presence or absence of a purine ring, a pyrimidine ring, or perhaps a substituent such as a keto-function in the correct place. Recognition of tRNA by the cognate synthetase would require all "yes" answers; one "no" would result in rejection by the enzyme. The difficulties encountered by researchers trying to elucidate the nature of tRNA recognition by synthetase demonstrate that the recognition process is a complex one (reviewed in ref. 161). The sequence data presented above are certainly insufficient to allow identification of all such determinants on $tRNAs^{Ser}$ from Drosophila. However, it may be possible to identify some of them. If we assume that the model for synthetase binding proposed by Rich and Schimmel (162), which suggests binding along the inside of the L-shaped tRNA molecule (Fig. 3), is correct, we can discard substantial blocks of shared sequence in the D and TψC loops (these are located on the outside of the L). If we further assume extensive conservation of recognition requirements between Drosophila and rat liver seryl-tRNA synthetases (which may well be true, since rat liver $tRNA_3^{Ser}$ is nearly identical to Drosophila $tRNA_{2b}^{Ser}$,

including modifications), we can consider the rat liver and Drosophila tRNAs^{Ser} together for homology. Within these constraints, numerous possibilities remain. Possible sites involved in specific synthetase recognition might include the D stem (especially the acetyl substituent in position 12), G at position 82 in the aminoacyl stem, and Um^{3}CU at the tip of the extra arm. All of these positions are conserved among six rat liver and Drosophila tRNA^{Ser} sequences.

I have largely avoided suggesting as possible recognition sites nucleosides located in stem structures (for example G30-C40 and A31-ψ39), or those that are likely by analogy with yeast tRNA^{Phe} to be involved in tertiary interactions necessary for compact tRNA folding (such as m^{5}C at position 57), though such nucleosides meet the other constraints imposed above.

It is doubtful that the positions I have suggested are sufficient for correct recognition. Heterologous aminoacylation experiments involving tRNAs^{Ser} and seryl-tRNA synthetases from yeast, Drosophila, and rat liver might help clarify this point.

C.3. Homologies among eukaryotic serine tRNAs; tRNA^{Ser} gene evolution

The nucleotide sequences of 10 eukaryotic serine tRNAs from S. cerevisiae, D. melanogaster, and rat liver were examined for sequence homologies (Fig. 13). Significant homology (≥ 66%) exists between any pair of tRNAs in this group. Indeed, 43 of 85 positions (51%) are conserved in all ten tRNAs, discounting differences in modification (Fig. 14). D. melanogaster tRNAs^{Ser} are generally less homologous to tRNAs^{Ser} from

% homology	Yeast minor (CGA)	Yeast minor (U*GA)	Yeast 1 (IGA)	Yeast 2 (IGA)	Rat liver 1 (IGA)	Rat liver 2a (CGA)	Rat liver 3 (GCU)	*Drosophila* 4 (CGA)	*Drosophila* 7 (IGA)	Drosophila 2b (GCU)
Yeast minor (CGA)[a]	--	96	85	88	76	86	71	69	66	69
Yeast minor (U*GA)[a]	--	--	85	88	76	85	71	70	68	69
Yeast 1 (IGA)	--	--	--	96	80	85	74	67	67	70
Yeast 2 (IGA)	--	--	--	--	79	84	73	68	68	71
Rat liver 1 (IGA)	--	--	--	--	--	96	76	76	78	74
Rat liver 2a (CGA)[b]	--	--	--	--	--	--	81	80	77	84
Rat liver 3 (GCU)	--	--	--	--	--	--	--	69	68	98
Drosophila 4 (CGA)	--	--	--	--	--	--	--	--	96	72
Drosophila 7 (IGA)	--	--	--	--	--	--	--	--	--	73
Drosophila 2b (GCU)	--	--	--	--	--	--	--	--	--	--

Figure 13 - Sequence homologies between eukaryotic serine tRNAs

a) Sequences for yeast minor serine tRNAs taken from ref. 67.

b) Complete sequence of 2a not determined. Comparison only for 70 positions determined for rat liver $tRNA^{Ser}_{2a}$ (ref. 176).

Figure 14 - Sequences of eukaryotic serine tRNAs

The nucleotide sequences of several serine tRNAs are shown, with base modifications indicated above the lines and ribose-methylations indicated below the lines. The sequences correspond to the following $tRNA^{Ser}$ species: 1, $tRNA^{Ser}_{minor}$ (CGA) from S. cerevisiae; 2, $tRNA^{Ser}_{minor}$ (U*GA) from S. cerevisiae; 3, $tRNA^{Ser}_{1}$ from S. cerevisiae; 4, $tRNA^{Ser}_{2}$ from S. cerevisiae; 5, $tRNA^{Ser}_{1}$ from rat liver; 6, $tRNA^{Ser}_{2a}$ from rat liver; 7, $tRNA^{Ser}_{3}$ from rat liver; 8, $tRNA^{Ser}_{2b}$ from D. melanogaster; 9, $tRNA^{Ser}_{4}$ from D. melanogaster; and 10, $tRNA^{Ser}_{7}$ from D. melanogaster. The lines beneath the D. melanogaster $tRNA^{Ser}_{7}$ sequence (#10) indicate positions where the base is conserved in all ten sequences (not counting modification differences). The nucleotide sequences in the figure were taken from ref. 67 (#1,2), ref. 176 (#6), Chapter III of this thesis (#8,9,10), or ref. 7 (#3,4,5,7).

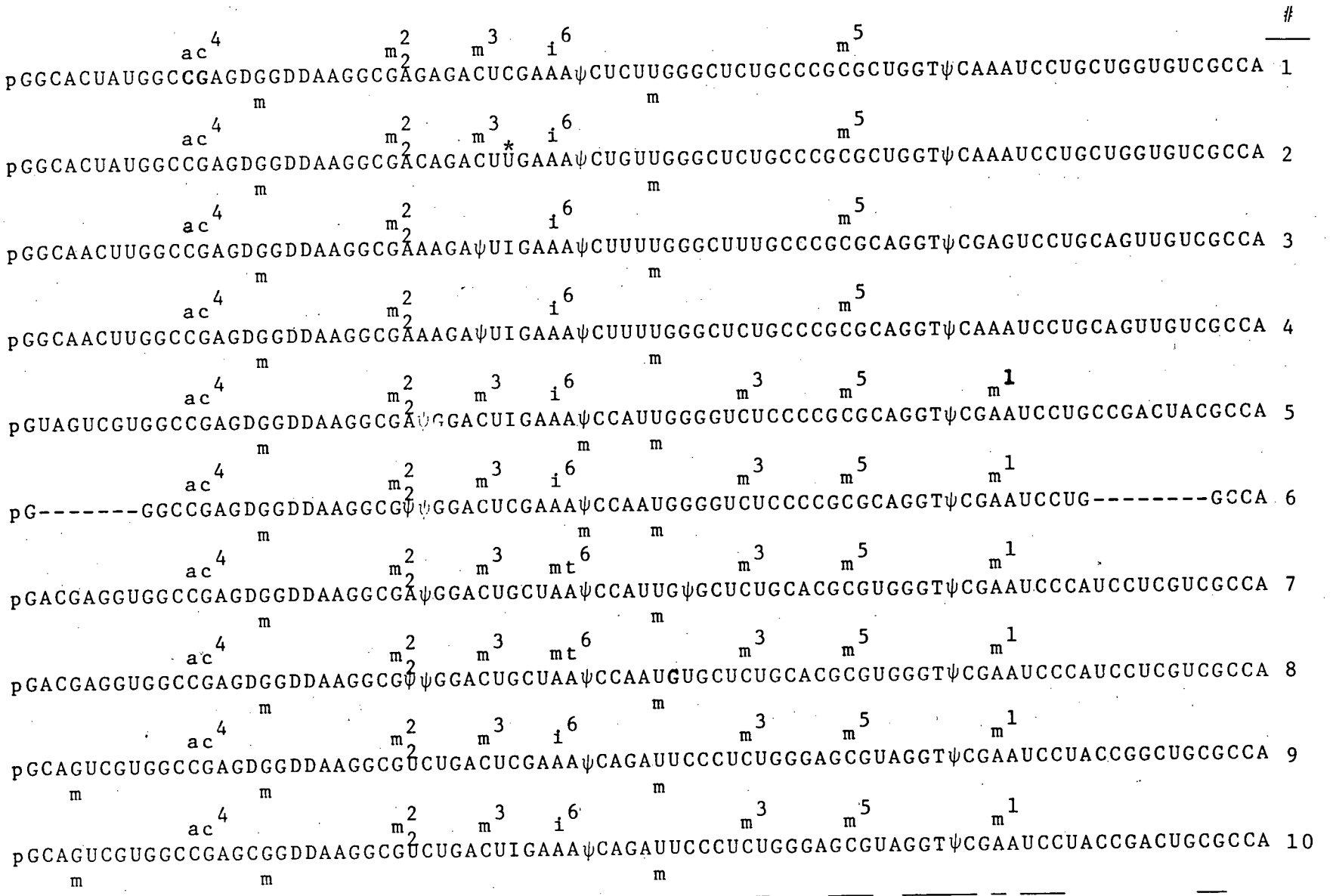

#
pGGCACUAUGGCCGAGDGGDDAAGGCGAGAGACUCGAAAψCUCUUGGGCUCUGCCCGCGCUGGTψCAAAUCCUGCUGGUGUCGCCA 1
pGGCACUAUGGCCGAGDGGDDAAGGCGACAGACUUGAAAψCUGUUGGGCUCUGCCCGCGCUGGTψCAAAUCCUGCUGGUGUCGCCA 2
pGGCAACUUGGCCGAGDGGDDAAGGCGAAAGAψUIGAAAψCUUUUGGGCUUUGCCCGCGCAGGTψCGAGUCCUGCAGUUGUCGCCA 3
pGGCAACUUGGCCGAGDGGDDAAGGCGAAAGAψUIGAAAψCUUUUGGGCUCUGCCCGCGCAGGTψCAAAUCCUGCAGUUGUCGCCA 4
pGUAGUCGUGGCCGAGDGGDDAAGGCGAψCGACUIGAAAψCCAUUGGGGUCUCCCCGCGCAGGTψCGAAUCCUGCCGACUACGCCA 5
pG-------GGCCGAGDGGDDAAGGCGψψGGACUCGAAAψCCAAUGGGGUCUCCCCGCGCAGGTψCGAAUCCUG--------GCCA 6
pGACGAGGUGGCCGAGDGGDDAAGGCGAψGGACUGCUAAψCCAUUGψGCUCUGCACGCGUGGGTψCGAAUCCCAUCCUCGUCGCCA 7
pGACGAGGUGGCCGAGDGGDDAAGGCGψψGGACUGCUAAψCCAAUGUGCUCUGCACGCGUGGGTψCGAAUCCCAUCCUCGUCGCCA 8
pGCAGUCGUGGCCGAGDGGDDAAGGCGUCUGACUCGAAAψCAGAUUCCCUCUGGGAGCGUAGGTψCGAAUCCUACCGGCUGCGCCA 9
pGCAGUCGUGGCCGAGCGGDDAAGGCGUCUGACUIGAAAψCAGAUUCCCUCUGGGAGCGUAGGTψCGAAUCCUACCGACUGCGCCA 10

FIGURE 14

the yeast S. cerevisiae than to those from rat liver. However, rat liver $tRNAs^{Ser}$ are equally homologous with Drosophila and yeast serine tRNAs. Unfortunately, the sequence for rat liver $tRNA^{Ser}_{2a}$ (anticodon CGA; ref. 176) is incomplete, while no sequence has been reported for a yeast $tRNA^{Ser}$ with anticodon GCU, limiting the value of such comparisons.

Nonetheless, it may be possible to draw limited conclusions at this point regarding the evolution of eukaryotic serine tRNA genes.

The $tRNAs^{Ser}$ with anticodons IGA and GCU from Drosophila and rat liver have been sequenced. Transfer RNA^{Ser}_1 (IGA) from rat liver is 76% homologous to $tRNA^{Ser}_3$ (GCU) from the same organism, differing at 20 of 85 positions. Similarly, the Drosophila tRNAs, $tRNA^{Ser}_7$ (IGA) and $tRNA^{Ser}_{2b}$ (GCU) are 73% homologous, differing at 23 of 85 positions. As shown in Fig. 15, the sites that differ between between $tRNAs^{Ser}$ (IGA and GCU) in these organisms are not distributed randomly. Sixteen of the 20 (rat liver) or 23 (D. melanogaster) differing sites in the tRNA "cloverleaf" coincide.

Particularly striking is the relationship seen with eukaryotic serine tRNAs specific for UCN codons. D. melanogaster $tRNA^{Ser}_4$ and $tRNA^{Ser}_7$ are 96% homologous, matching at 82 of 85 positions. One of the three differences is in first position of the anticodon (the "wobble" position). Four serine tRNAs from yeast reading UCN codons have been sequenced: $tRNA^{Ser}_1$ (anticodon IGA), $tRNA^{Ser}_2$ (IGA), $tRNA^{Ser}_{UCG}$ (CGA), and $tRNA^{Ser}_{UCA}$ (U*GA). Transfer RNA^{Ser}_2, the predominant species with anticodon IGA, is 88% homologous to both $tRNA^{Ser}_{UCG}$ and $tRNA^{Ser}_{UCA}$. The latter two tRNAs are 96% homologous. They differ at only three positions, one being the "wobble" position of the anticodon. (The other two changes are the result of a G-C to C-G

Figure 15 - Comparison of positions differing between serine tRNAs with anticodons IGA and GCU in Drosophila and rat liver.

The cloverleaf structure of $tRNA_{2b}^{Ser}$ from Drosophila melanogaster (which differs from rat liver $tRNA_3^{Ser}$ by a single base pair) is shown. Solid lines adjacent to the cloverleaf structure indicate positions differing between Drosophila $tRNA_{2b}^{Ser}$ (GCU) and $tRNA_7^{Ser}$ (IGA). Wavy lines indicate positions differing between rat liver $tRNA_3^{Ser}$ (GCU) and $tRNA_1^{Ser}$ (IGA).

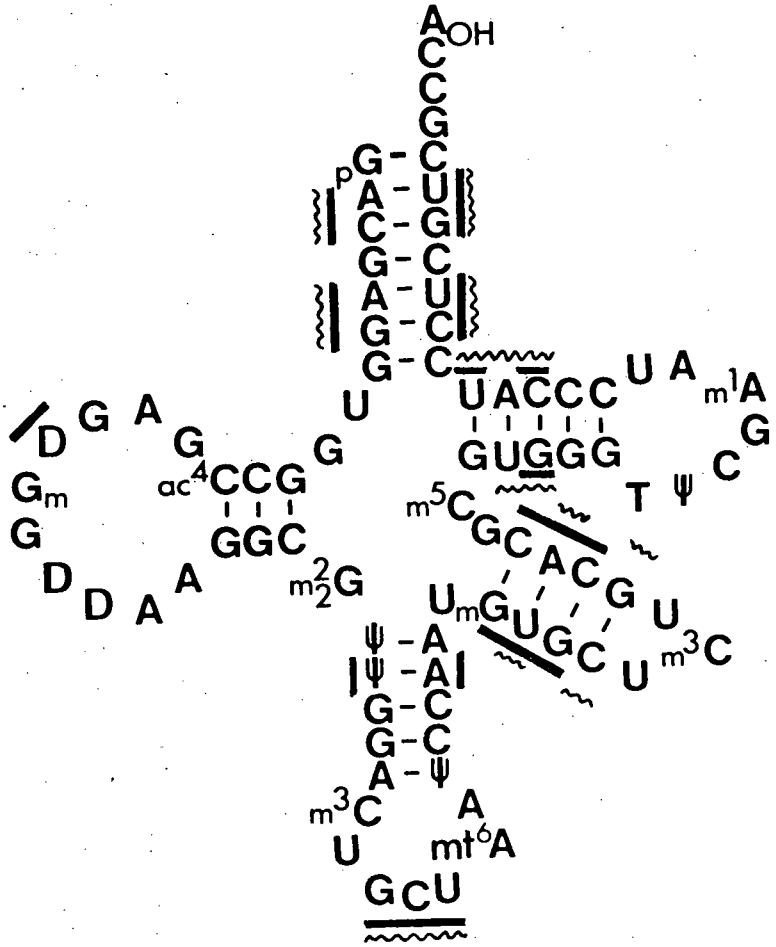


Figure 15.

transversion in the anticodon stem.) The partial sequence of rat liver $tRNA_{2a}^{Ser}$ (CGA) differs from the completed sequence of rat liver $tRNA_{1}^{Ser}$ (IGA) at 3 of 70 corresponding positions. As with the yeast tRNAs mentioned above, two of the differences are the result of a base pair transversion in the anticodon stem (ψ-A to A-U) while the third is in the "wobble" position of the anticodon. Though $tRNAs^{Ser}$ with the same anticodon (IGA or CGA) from these organisms differ by up to 32% (Fig. 13), $tRNAs^{Ser}$ with different anticodons within each species are nearly identical. This high extent of homology cannot be required for serine tRNAs in these organisms to function in protein synthesis, since serine tRNAs reading AGPy codons in rat liver and D. melanogaster differ from those reading UCN codons by up to 30%. The existence of these very similar $tRNAs^{Ser}$ with different specificities raises the question of how the integrity of reiterated genes is maintained in eukaryotes.

One interpretation of these data is that the genes for eukaryotic $tRNAs^{Ser}$ specific for UCN codons are in some sense evolutionarily "linked", and cannot evolve independently. (Such an interpretation is relevant to the discussion of Chapter IV, dealing with $tRNA^{Ser}$ in D. melanogaster).

It is curious, and may well be significant, that the very similar pairs of $tRNAs^{Ser}$ in Drosophila and rat liver (anticodons IGA and CGA) are not equivalent to the pair in yeast (U*GA and CGA). This suggests that the evolution of eukaryotic serine tRNA genes has been quite a complex process. It must be concluded here that too few eukaryotic $tRNA^{Ser}$ sequences are known to allow extensive conclusions to be drawn concerning evolution of their genes.

Chapter IV

STRUCTURE AND ORGANIZATION OF SERINE tRNA GENES IN DROSOPHILA

A. Results

Dunn et al. (114) isolated seven recombinant plasmids containing D. melanogaster DNA that hybridize $tRNA^{Ser}_{4,7}$. Three of these (pDt 1, pDt 5, and pDt 81) were shown by restriction analysis and Southern blotting experiments (114) to contain the same insert. By in situ hybridization of radiolabelled plasmid to polytene chromosomes, pDt 5 was shown to derive from the minor $tRNA^{Ser}_{4,7}$ site at 23E on chromosome 2L (114). The four remaining plasmids - pDt 16, pDt 17R, pDt 27R, and pDt 73 - all derive from the 12DE region of the X chromosome (114,155,156). These plasmids were chosen for further study.

A.1. Results of Southern blot analysis of recombinant plasmids using tetramer-recognizing restriction enzymes

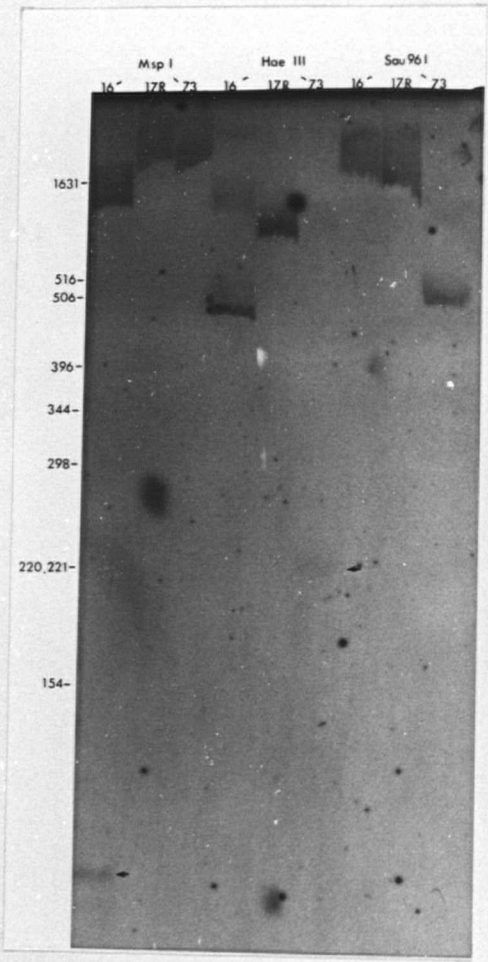
The DNAs of pDt 16, pDt 17R and pDt 73 were cleaved by the restriction enzymes Hae III ($GG^{\downarrow}CC$), Msp I ($C^{\downarrow}CGG$), and Sau96 I ($G^{\downarrow}GNCC$), separated according to size on a 5% polyacrylamide gel, denatured and blotted onto nitrocellulose filters (150), then hybridized with [3'-^{32}P] $tRNA_7^{Ser}$. The resulting pattern is shown in Fig. 16a. The DNAs of pDt 27R, pDt 17R, and pDt 73 were cleaved by the restriction enzymes Taq I

Figure 16 - Southern blot analysis of recombinant plasmids hybridizing *Drosophila* serine tRNAs.

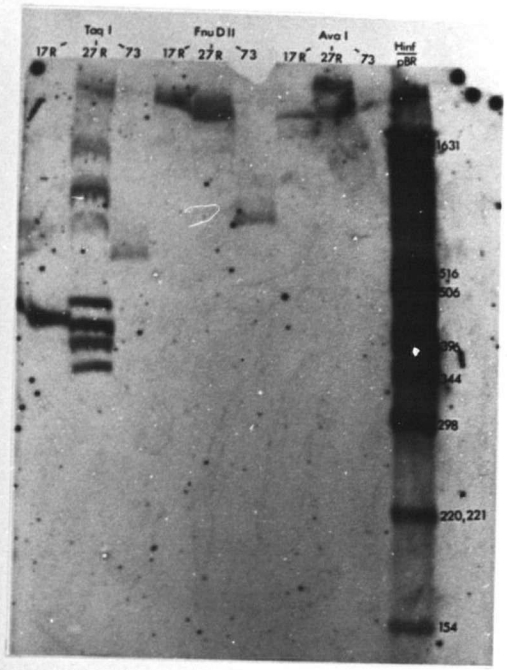
(a) Plasmid DNAs (1μg/slot) were digested with *Msp* I, *Hae* III, or *Sau*96 I. The resulting fragments were separated by electrophoresis on a non-denaturing 5% polyacrylamide gel for 16 hr. at 100V, transferred to a nitrocellulose membrane, hybridized with [3'-^{32}P]tRNA$_7^{Ser}$ and washed as described in Chapter II. Autoradiography was for 5.5 days at -70° using activated "screen" X-ray film (as described in Methods, ref. 43).

(b) Plasmid DNAs (1 μg/slot pDt 17R and pDt 73, 2 μg/slot pDt 27R) were digested with *Taq* I, *Fnu*D II, or *Ava* I. The resulting fragments were treated as in (a). Autoradiography was for 53 hr. at -70° using activated Kodak XR-1 screen X-ray film (Methods, ref. 43).

a


Msp I
Hae III
Sau 96 I
16´ 17R ´73
1631-
516-
506-
396-
344-
298-
220,221-
154-

b


Taq I
Fnu D II
Ava I
Minf pBR
17R´ 27R ´73
1631
516
506
344
298
220,221
154

($T^{\downarrow}CGA$), FnuD II ($C^{\downarrow}GCG$), and Ava I ($C^{\downarrow}PyCGPuG$) and analyzed as above (Fig. 16b).

From these experiments, several points became clear. First, since Hae III and Taq I are both expected to cut within $tRNA^{Ser}_{4,7}$ genes based on the tRNA sequences (Fig. 17), it is possible to estimate the number of genes present on each plasmid. pDt 17R and pDt 73 both contain a single $tRNA^{Ser}$ gene, as a single $tRNA^{Ser}_{7}$-hybridizing fragment was generated by each of the six restriction enzymes used to cleave these plasmids (Fig. 16). pDt 16 contains two or more $tRNA^{Ser}$ genes, since cleavage with Hae III and Msp I results in two hybridizing fragments (Fig. 16a). Cleavage of pDt 16 with Sau96 I produces a single hybridizing fragment approximately 1.2 kb in length. Thus, the two $tRNA^{Ser}$ genes are located fairly close together. Cleavage of pDt 27R DNA with Taq I and Southern blot analysis showed four hybridizing fragments in this plasmid with lengths of about 345 b.p., 370 b.p., 395 b.p., and 435 b.p. Taq I is expected to cleave $tRNA^{Ser}_{4}$ genes in two places, $tRNA^{Ser}_{7}$ genes in one (Fig. 17). Conceivably, cleavage at one Taq I site in a $tRNA^{Ser}_{4}$ gene might inhibit cleavage at the second adjacent site in the gene. If so, partial cleavage of a single $tRNA^{Ser}_{4}$ gene might result in three $tRNA^{Ser}$-hybridizing fragments: two extending from the first Taq I site 5' to the gene to the two Taq I sites in the gene at positions 33-36 and 64-67 (the numbering is based on the $tRNA^{Ser}_{4,7}$ sequences, Fig. 8), and a third extending from the Taq I site at positions 33-36 of the $tRNA^{Ser}_{4}$ sequences to the first restriction site 3' to the gene. (The fragment extending from position 65 to the first Taq I site 3' to the gene would not form stable hybrids with $tRNA^{Ser}$, as shown by the single hybridizing band produced by cleavage of pDt 17R.) The fourth band would arise from a $tRNA^{Ser}_{7}$ gene yielding a single hybridizing

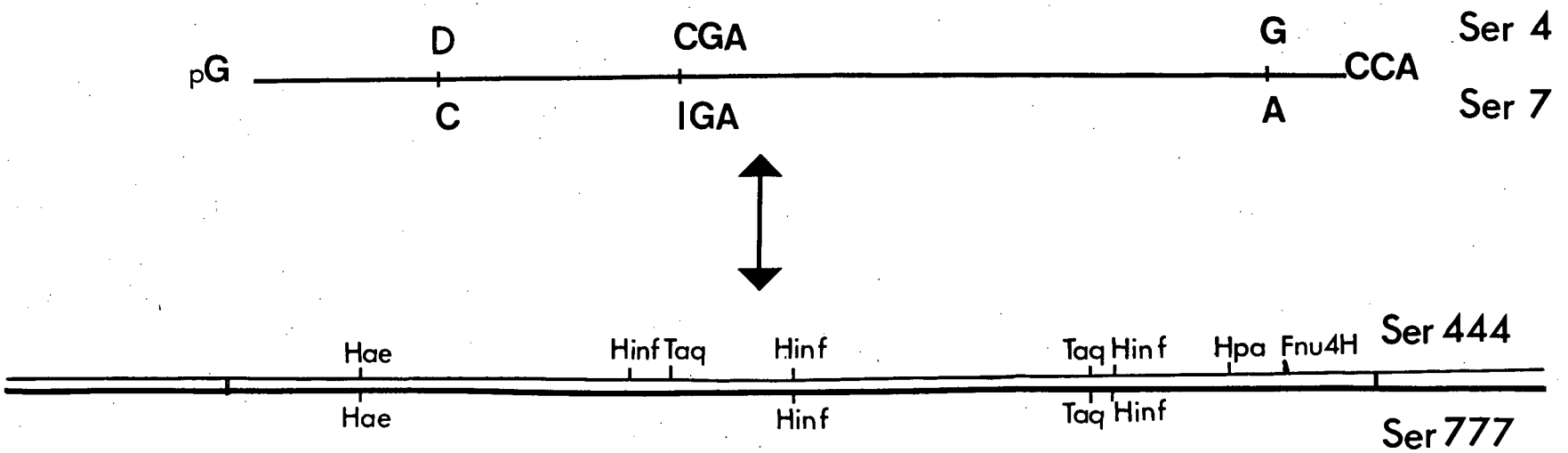


Figure 17 - Restriction maps expected for $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ genes.

Taq I fragment. Cleavage of pDt 27R with Ava I gave two large hybridizing fragments, supporting the interpretation that this plasmid contains two or more $tRNA^{Ser}$ sequences. Digestion with Taq I was not complete, as shown by the presence of several light bands in addition to the four strong bands on the autoradiograph (Fig. 16b). However, the roughly equal intensities of the four hybridizing bands resulting from Taq I cleavage of pDt 27R and the adjacent single band resulting from cleavage of a roughly equivalent amount of pDt 17R DNA suggests that pDt 27R does contain four $tRNA^{Ser}$ genes. In the absence of direct evidence to the contrary, this simplest interpretation of the data is the preferred one.

The data shown in Figures 16a and 16b therefore indicate that the four plasmids contain a total of eight $tRNA^{Ser}$ genes (one each in pDt 17R and pDt 73, two in pDt 16, and four in pDt 27R). As none of the restriction patterns observed for these plasmids appear similar, there seems to be no reason to suspect that any pair of the four plasmids might be allelic. Therefore, the 12DE region of the D. melanogaster genome containing $tRNA^{Ser}$ genes includes at least 19 kb of DNA sequence (encompassing the inserts of the four plasmids examined) and at least eight $tRNA^{Ser}$ genes.

This interpretation is consistent with results of in situ hybridization of $tRNA^{Ser}_{4,7}$ to polytene chromosomes. Based on relative numbers of silver grains observed at different chromosomal regions, it can be estimated that the 12DE region of the X chromosome contains roughly 5-10 times more $tRNA^{Ser}$ genes than the minor autosomal sites (109). Assuming one or two genes at the minor sites, the grain numbers are consistent with roughly 10 to 20 genes at the 12DE site.

A.2. Strategy for DNA sequence analysis of recombinant plasmids

The strategy for DNA sequence analysis of recombinant plasmids consisted of four steps. (i) The plasmid was mapped with restriction enzymes, and tRNA genes located in the plasmid by Southern blotting analysis. (ii) Based on the map deduced, a DNA fragment containing the gene(s) was purified for sequence analysis. (iii) The restriction enzymes known from sequence analysis of $tRNA^{Ser}_{4,7}$ to cut within $tRNA^{Ser}$ genes were used to obtain the gene sequence and some flanking sequences by the method of Maxam and Gilbert (152). (iv) Restriction enzyme sites located outside the genes were used to sequence back into the gene and verify the restriction sites inside the gene used in step (iii).

A.3. Sequence analysis of pDt 73

Southern analysis of pDt 73 showed that the single $tRNA^{Ser}$ gene was located within a 520 b.p. Sau96 I fragment which, when isolated after [3'-^{32}P]-labelling with E. coli DNA polymerase I (Klenow fragment) and [α-^{32}P] dGTP, was found to contain single Hae III and Taq I restriction sites roughly in the middle. This fragment was purified for further sequence analysis by preparative polyacrylamide gel electrophoresis. It was expected to contain a single gene plus about 200 b.p. of 5' and 3' flanking sequences for comparison with other $tRNA^{Ser}$ genes from 12DE and from autosomal sites.

Various restriction endonuclease sites in the 520 b.p. Sau96 I fragment were specifically 3'-end labelled. This was accomplished by by using the specificity of E. coli DNA polymerase I to label either the

Sau96 I ends (5'-extended, G$^{\downarrow}$GNCC) and separate the labelled ends with a second restriction enzyme; or to label the ends generated by cleavage with a second enzyme. For example, cleavage of the 520 b.p. Sau96 I fragment with Taq I (T$^{\downarrow}$CGA) allowed exclusive labelling of the Taq I ends with [α-^{32}P] dCTP and E. coli DNA polymerase I (Klenow fragment). The two resulting single end-labelled fragments were purified by polyacrylamide gel electrophoresis and subjected to Maxam-Gilbert chemical modification reactions. In the complementary experiment, the 520 b.p. fragment was cleaved by Taq I, labelled at the Sau96 I ends with [α-^{32}P] dGTP by E. coli DNA polymerase I (Klenow fragment) gel-purified and sequenced. The nucleotide sequence at the Taq I restriction site (in the gene) was verified by sequencing from the Hind III site just 3' to the gene. Specific labelling of the Hind III site (A$^{\downarrow}$AGCTT) with [α-^{32}P] dATP by E. coli DNA polymerase I again allowed the one-step purification by polyacrylamide gel electrophoresis of a single-end labelled substrate for the sequencing reactions.

Note that as the recombinant plasmids contain Hind III-cut D. melanogaster DNA, the Hind III site adjacent to the tRNA^{Ser} gene of pDt 73 marks the junction between insert and pBR 322 DNA sequences. Thus, only eleven nucleotides of 3'-flanking sequence could be determined for this tRNA^{Ser} gene.

All nucleotides in the sequence shown in Fig. 18 were obtained in at least two independent experiments except 139-167, which were only obtained once. Due to the scarcity of restriction sites in the very A-T rich 5'-flanking region, little of the sequence was obtained from both strands. In regions where the sequences of both strands were determined

Figure 18 - Nucleotide sequence of the tRNA gene-containing region in pDt 73.

(a) The compiled sequence of the gene-containing region is shown (non-coding strand). The 474 gene sequence is underlined.

(b) A restriction map of the gene-containing region in pDt 73, indicating the extents and directions of sequences obtained in Maxam/Gilbert sequencing experiments, is shown. Several relevant restriction enzyme cleavage sites are indicated: S (Sau96 I), Hd (Hind III), T (Taq I), and Hf (Hinf I).

a. The Nucleotide Sequence of pDt73

```
        10         20         30         40         50         60         70         80         90        100
TTACGCAAAT TTAATTGGAC ATAGATAGAT AACGCGTGAG TTTAAGTTCA ATTCTTATCT ATTAAGTATT TCGTTTAGGC GCATTCCAAA GTATTTATAT

       110        120        130        140        150        160        170        180        190        200
TTAGTTATCA GTTCCGAAAT TCCAATTTAT ATTATCAGCT TAGGTTTTGC ACAAGATATG GAAAATATTT TTTGTTTTTG TAAATTAATA TAATACTATT

       210        220        230        240        250        260        270        280        290        300
AACTTTATAT TACTTTCTTA AATTTTATTG ATATTTTTTG CGCATATATC AAGGCAGTCG TGGCCGAGTG GTTAAGGCGT CTGACTAGAA ATCAGATTCC

       310        320        330        340
CTCTGGGAGC GTAGGTTCGA ATCCTACCGG CTGCGCTTGT AAGCTT
```

b.

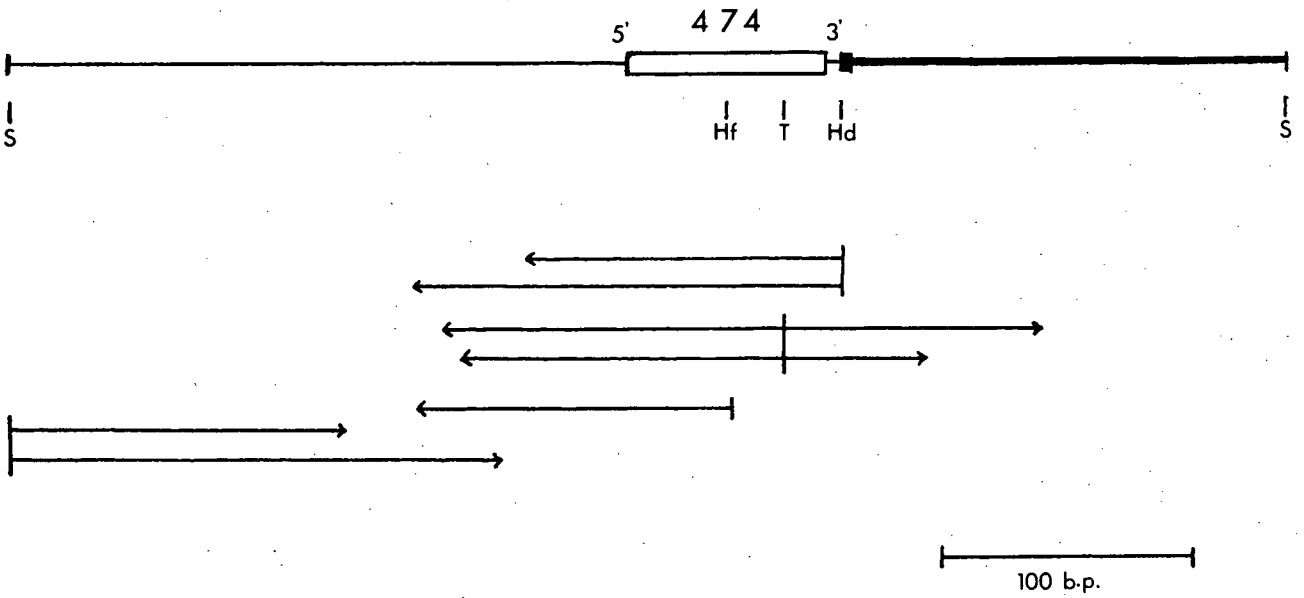
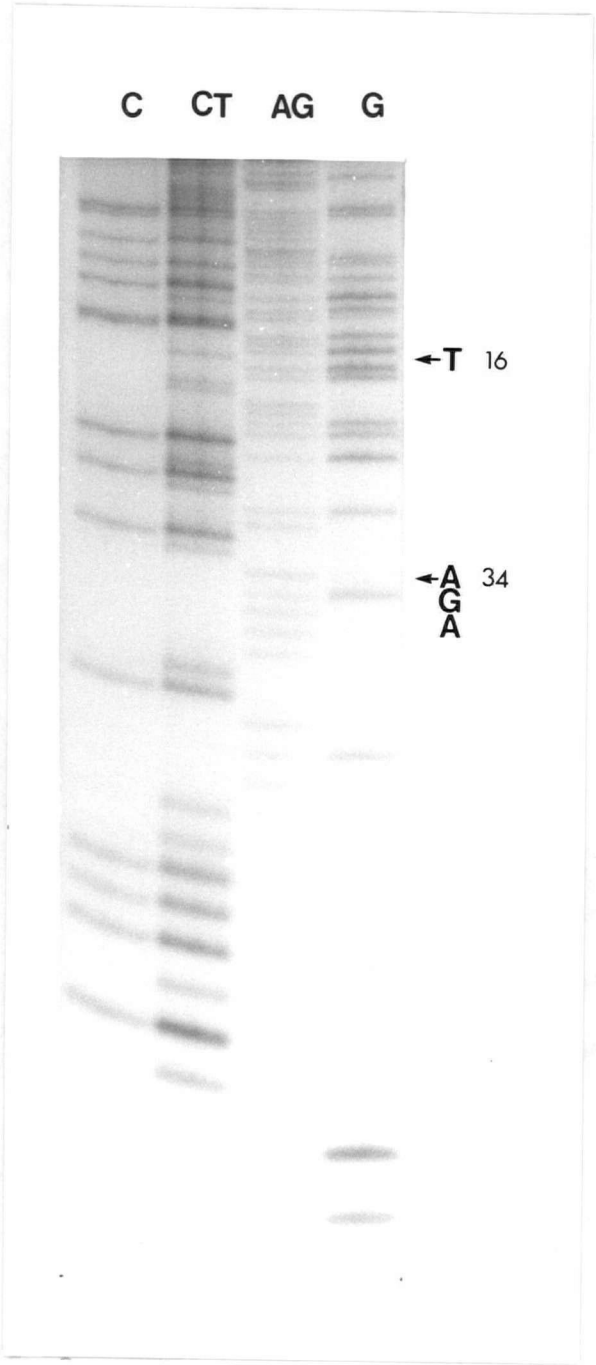


Figure 19 - Maxam-Gilbert sequence analysis of the serine tRNA gene in pDt 73.

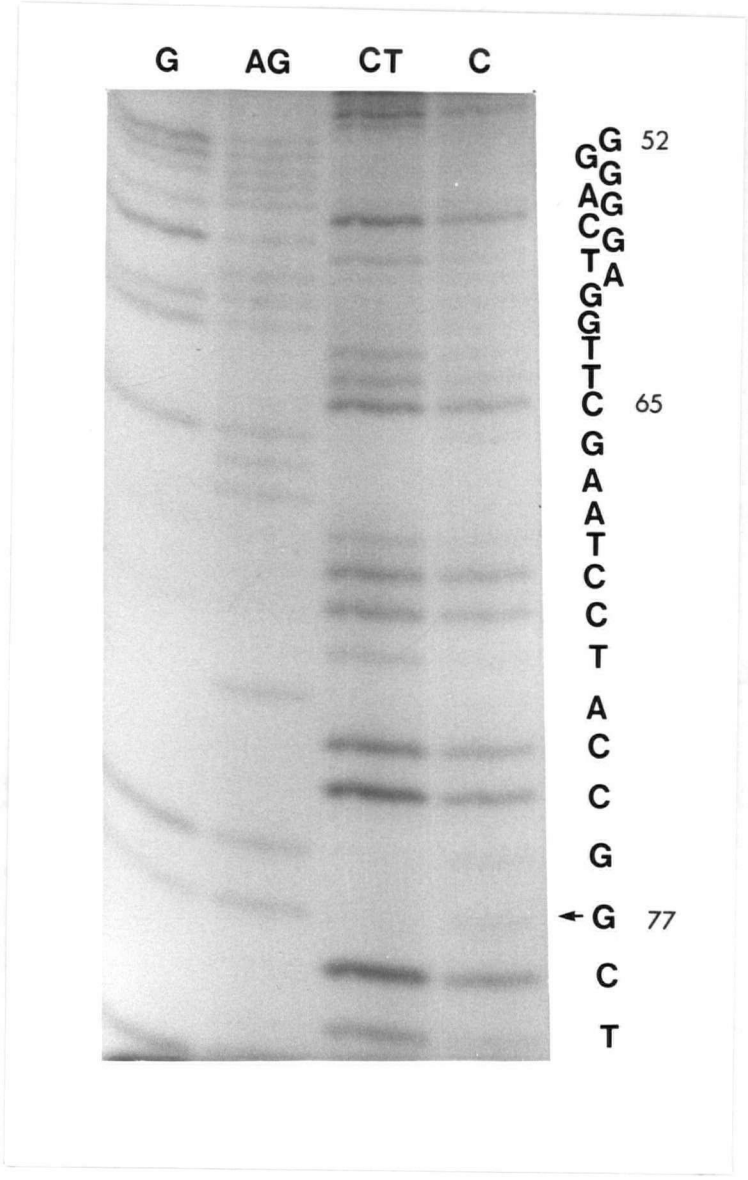
(a) Sequence analysis of a Taq I/Sau96 I restriction fragment labelled at the Taq I end with [α-^{32}P]dCTP and E. coli DNA polymerase I (large fragment). The sequence shown is on the non-coding strand. Positions distinguishing $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ are indicated with arrows. Numbers indicate the positions in the $tRNA^{Ser}$ sequence (Figure 8).

(b) Sequence analysis of a Hind III/Sau96 I restriction fragment labelled at the Hind III end with [α-^{32}P]dATP and E. coli DNA polymerase I (large fragment). The sequence shown is on the non-coding strand. Position 77 is one of three distinguishing $tRNA_4^{Ser}$ from $tRNA_7^{Ser}$ (Figure 8).

a.


C
CT
AG
G
T 16
A 34
G
A

b.


G
AG
CT
C
G 52
G
G
A
C
T
G
T
T
C 65
G
A
A
T
C
C
T
A
C
C
G
G 77
C
T

the agreement was very good (98-100%). The $tRNA^{Ser}$ gene sequence was obtained in multiple experiments. The only ambiguous position encountered in the gene sequence was the G at position 10 of the $tRNA^{Ser}_{4,7}$ sequence, where a light band was consistently encountered in the G-slot of autoradiographs. This position was confirmed as G by cleavage of a Hinf I (^{32}P-dAMP)/Sau96 I fragment with Hae III (which should cleave only if G is present at position 10) and electrophoresis of the digestion product alongside the products of Maxam-Gilbert sequencing reactions (results not shown).

The nucleotide sequence of the $tRNA^{Ser}$ gene in pDt 73 does not correspond to either $tRNA^{Ser}_4$ or $tRNA^{Ser}_7$. Genes for those tRNAs can be represented as 444 or 777 genes depending on the nucleotides present at positions +16, +34, or +77 of the tRNA sequences (the three differing positions). A $tRNA^{Ser}_4$ gene (444) should contain T-16, C-34, and G-77, while a $tRNA^{Ser}_7$ gene (777) should contain C-16, A-34, and A-77. The $tRNA^{Ser}$ gene in pDt 73 is a "hybrid" 474 gene. It contains an A in the anticodon "wobble" position (+34) corresponding to inosine in $tRNA^{Ser}_7$, but T-16 and G-77 corresponding to D and G at these positions in $tRNA^{Ser}_4$. As shown in Fig. 17, two of the three nucleotide changes distinguishing $tRNA^{Ser}_4$ from $tRNA^{Ser}_7$ should result in the presence of restriction sites in presumptive 444 genes not found in 777 genes. The restriction patterns observed from the 520 b.p. fragment examined are in agreement with the presence of a mixed 474 sequence (lacking the 3'-CCA of the mature tRNA). Examples of Maxam/Gilbert sequencing experiments are shown in Figure 19.

A.4. Sequence analysis of pDt 17R

The restriction enzyme map of pDt 17R was determined by agarose gel electrophoretic analysis of DNA fragments produced by digestion with one or two restriction enzymes recognizing hexanucleotide sequences (for example, Fig. 20a). This map was confirmed by Southern blotting experiments to locate fragments containing sequences homologous to a [3'-^{32}P] $tRNA_7^{Ser}$ probe (for example, Fig. 20b). The tRNA-hybridizing sequence was localized to a 1.5 kb segment of the 3.3 kb pDt 17R insert bounded by a Pvu II site at the left end and an Xho I site at the right end (as drawn in Fig. 21). The completed restriction map for the insert is presented in Fig. 21.

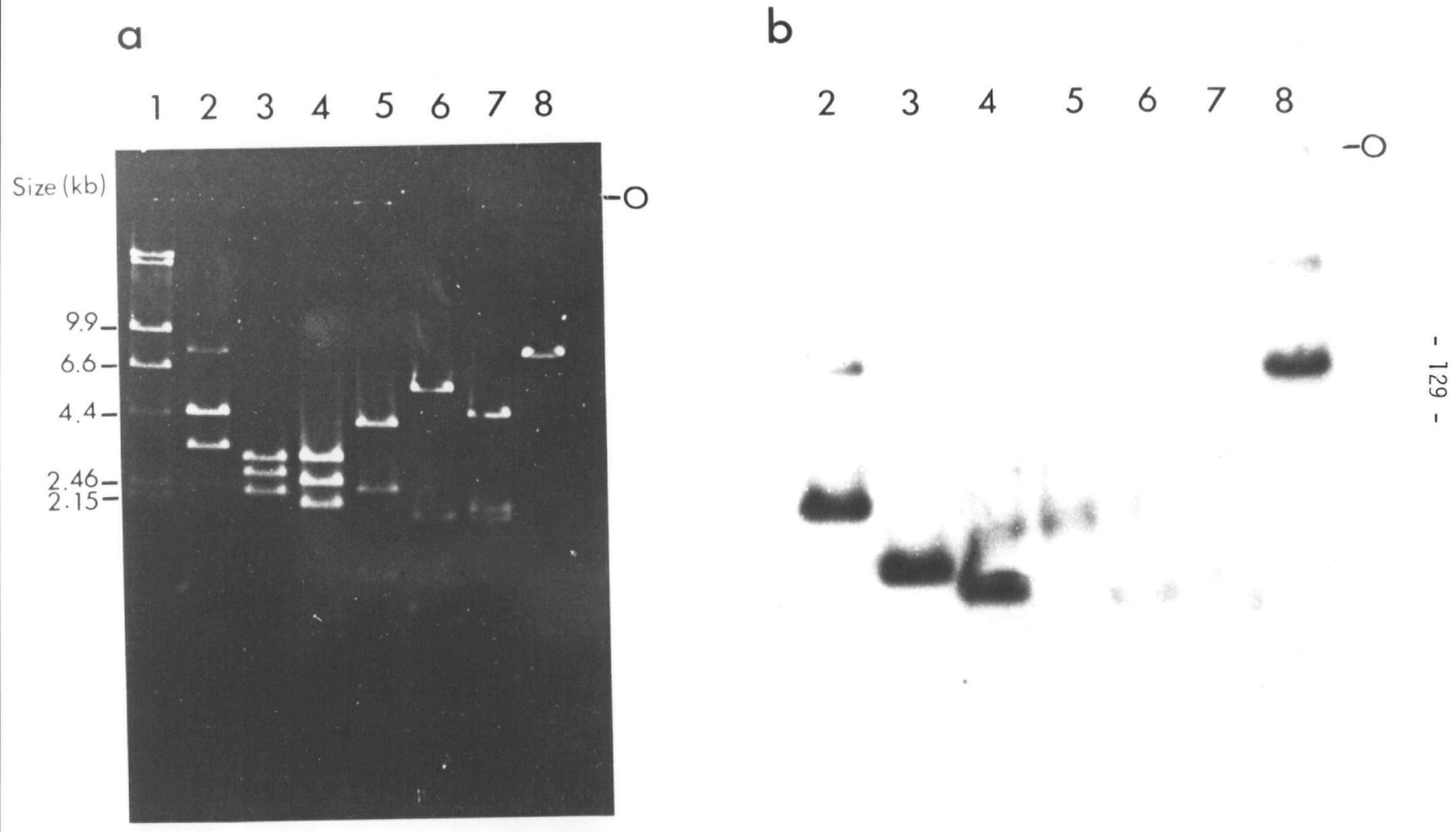
The DNA segment isolated for sequence analysis was the 1.75 kb EcoR I/Xho I fragment containing the left half of the insert (Fig. 21). It was chosen because it migrated well away from other digestion products on 0.5% agarose gels (Fig. 20a) and could thus be readily purified in bulk.

Fine structure mapping of the 1.75 kb EcoR I/Xho I fragment was accomplished in two ways. First, the set of radiolabelled fragments derived from the 1.75 kb fragment by digestion with Taq I and 3'-end labelling was digested with a number of other restriction enzymes, the various fragments separated by polyacrylamide gel electrophoresis and located by autoradiography. Such an experiment yields considerable information regarding possible restriction enzymes appropriate for secondary cleavage of double-end labelled fragments in Maxam-Gilbert sequence analysis, as well as helping to identify tRNA gene-containing fragments and providing

Figure 20 - Restriction mapping of pDt 17R.

(a) DNA was digested with Hind III (slot 2), Pvu II (3), Pvu II and EcoRI (4), EcoRI and BamHI (5), EcoRI and Xho I (6), Hind III and Xho I (7), or Pst I (8), and the resulting fragments separated by electrophoresis on a 0.5% agarose gel as described in Chapter II. Size standards were Hind III fragments of bacteriophage λ DNA (slot 1).

(b) Restriction fragments from the gel shown in part (a) were transferred to nitrocellulose sheets and hybridized with [3'-^{32}P]tRNA$_7^{Ser}$ (Southern blotting; Chapter II). Autoradiography was for 5 days.


a
1 2 3 4 5 6 7 8
Size (kb)
9.9
6.6
4.4
2.46
2.15
-O
b
2 3 4 5 6 7 8
-O

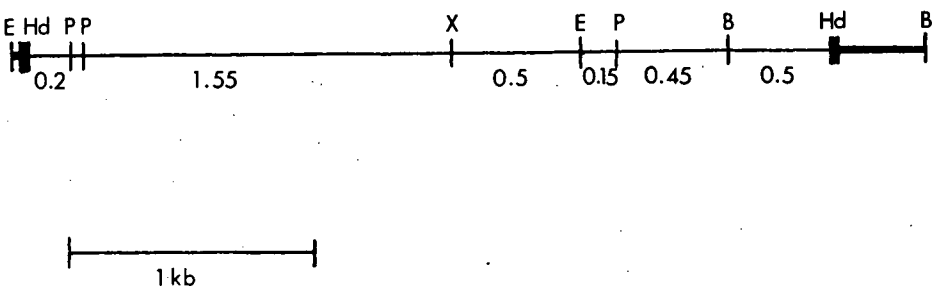


Figure 21 - Restriction enzyme map of the Drosophila insert in pDt 17R.

Abbreviations used for restriction enzymes are E (EcoRI), Hd (Hind III), P (Pvu II), X (Xho I), and B (BamHI). Heavy lines at each end represent pBR 322 sequences.

some evidence regarding the relative positions of the Taq I fragments. Second, mapping of the 1.75 kb fragment by the method of Smith and Birnstiel (151) after specifically labelling the EcoR I site using E. coli DNA polymerase I and [α-^{32}P]dATP allowed the tRNA^{Ser} gene to be located and its orientation relative to the EcoR I site to be established (Fig. 22). This procedure also provided useful information regarding appropriate second enzymes for cleaving double-end labelled restriction fragments in Maxam-Gilbert DNA sequence analysis.

Based initially on the map information above, and drawing on DNA sequence data as they became available, a continuous nucleotide sequence was obtained containing the tRNA^{Ser} gene plus about 250 nucleotides of 5'-flanking sequence and 210 nucleotides of 3'-flanking sequence. The extents and directions of sequences obtained in various experiments are shown in Fig. 23b. The nucleotide sequence of this region is presented in Fig. 23a. The presence of more restriction sites in the flanking sequences than found for pDt 73 allowed most of the 540 base pair sequence to be determined on both strands. The error frequency in sequence determination is estimated to be less than 1%. The tRNA^{Ser} gene sequence was unambiguous. A sample experiment is shown in Fig. 24.

The single tRNA^{Ser} gene present in pDt 17R is a 777 gene corresponding exactly to the tRNA$_7^{Ser}$ sequence, except for the 3'-terminal CCA of the tRNA which is not gene-coded.

A.5. Sequence analysis of pDt 16

The Southern blotting experiments described in section A.1. above indicated the presence of a very small (~ 100 b.p.) tRNA^{Ser}-hybridizing

Figure 22 - Smith/Birnstiel restriction mapping of the gene-containing region in pDt 17R.

Mapping of a 1.75 kb EcoRI/Xho I restriction fragment labelled at the EcoRI site with [α-^{32}P]dATP and E. coli DNA polymerase I was accomplished using the method of Smith and Birnstiel (Chapter II). The fragment was partially digested with Hae III (Hae), Hinf I (Hinf), Taq I (Taq), Msp I (Msp), Hha I (Hha), Sau96 I (Sau96), or Pvu II (Pvu). The size standards were Hinf I-derived fragments of pBR 322 labelled with [α-^{32}P]dATP and E. coli DNA polymerase I (pBR/Hinf). Electrophoresis was for 15 hr. at 150V on a 5% polyacrylamide gel. Each gel slot contained 25,000 Cerenkov cpm. Autoradiography was for 38 hr. at -20° using Kodak XR-1 "screen" X-ray film.

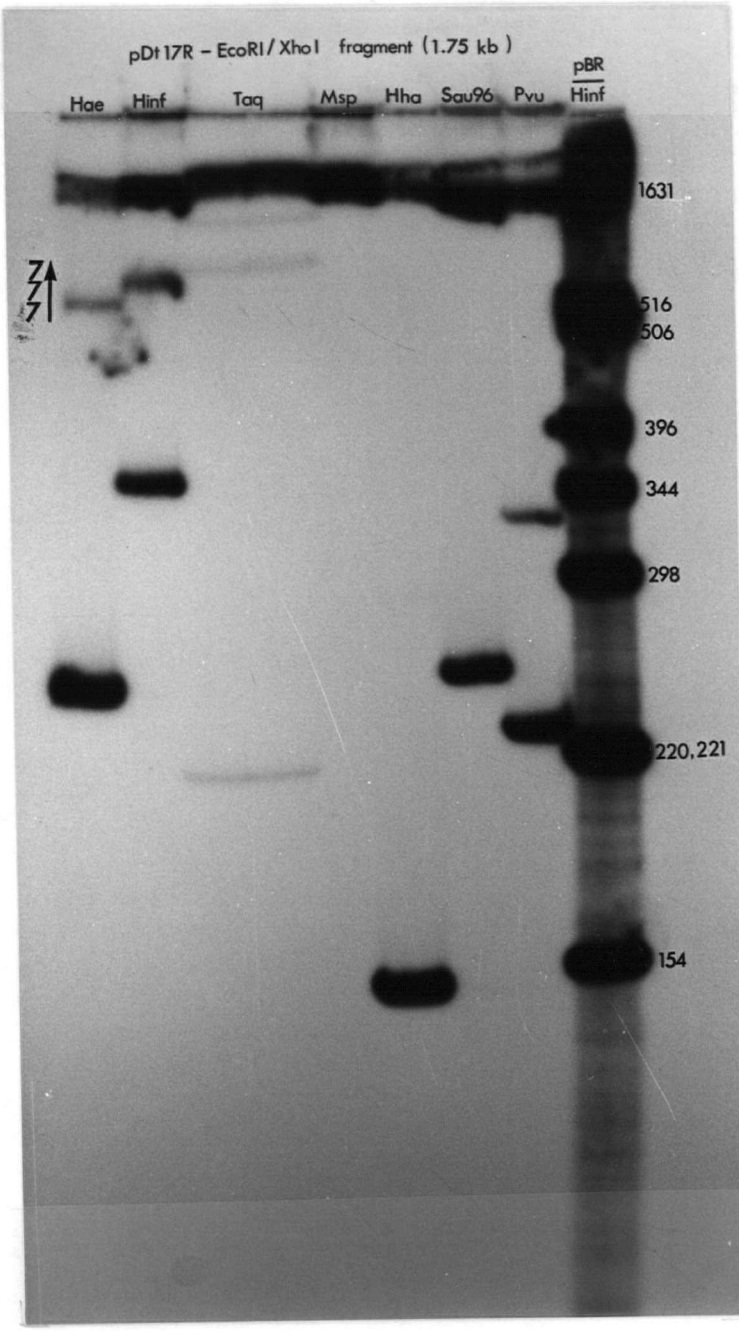

pDt 17R – EcoRI / Xho I fragment (1.75 kb)
Hae
Hinf
Taq
Msp
Hha
Sau96
Pvu
pBR
Hinf
1631
516
506
396
344
298
220,221
154

Figure 23 - Nucleotide sequence of the tRNA gene-containing region in pDt 17R.

(a) The compiled sequence of the gene-containing region is shown (non-coding strand). The 777 gene sequence is underlined.

(b) A restriction map of the gene-containing region in pDt 17R, indicating the extents and directions of sequences obtained in Maxam/Gilbert sequencing experiments, is shown. A number of relevant restriction enzyme cleavage sites are indicated: P (Pvu II), S (Sau96 I), Hf (Hinf I), Ha (Hae III), T (Taq I), and S3 (Sau3A I).

a. The Nucleotide Sequence of pDt17R

```
        10         20         30         40         50         60         70         80         90        100
CAGCTGAAGG GATGAACAGA GGGGGGCCTT GCGTTCCCAC TTGAGTGGCA AACAAATAGA TTAGAAATAG ATTTCCCAAA GCGAGAAAGT TTATTTCAAC

       110        120        130        140        150        160        170        180        190        200
AGCTGTGGGT GCTAAAGATA ACATACTGCA TTTCTAAAGG TTGAGTCAAT AAATCTGTCA TGAATATTTT AATTAACTGT GGTTTCATAA AGCGAACGGA

       210        220        230        240        250        260        270        280        290        300
AAGAAACCAA GAAACTTCCA GCTATAAATA AAAAAAGCTG AACAAATAGC GTATTTAAAT AGCTAACTAA AGTATCTATC AGATAGTATC TGCGACCCAA

       310        320        330        340        350        360        370        380        390        400
CCTCTTGCAC CTCTTGAGAA CTCAATTTTC GCCACCCACC CATCAAGCAG TCGTGGCCGA GCGGTTAAGG CGTCTGACTA GAAATCAGAT TCCCTCTGGG

       410        420        430        440        450        460        470        480        490        500
AGCGTAGGTT CGAATCCTAC CGACTGCGAT ATGAAGAGTA TCTTTTTTAT GTCAGATACT TTTATGTATC TATGGGATCA ACGATCTTAA AGATATACAC

       510        520        530        540        550        560        570        580        590        600
ATAATTCAAT AATTGTCTAT GGCTTAAGTA GCCTATTGGT TAGGTTGTAC GCTTATGAAA CCGCAGTTTC GAGTTCAATT CAATGCGTAG TAGTGCTTTA

       610        620        630
TTTATAATGA ACTGTTTGTT AGTAGTTCGT AATTTTTA
```


b.

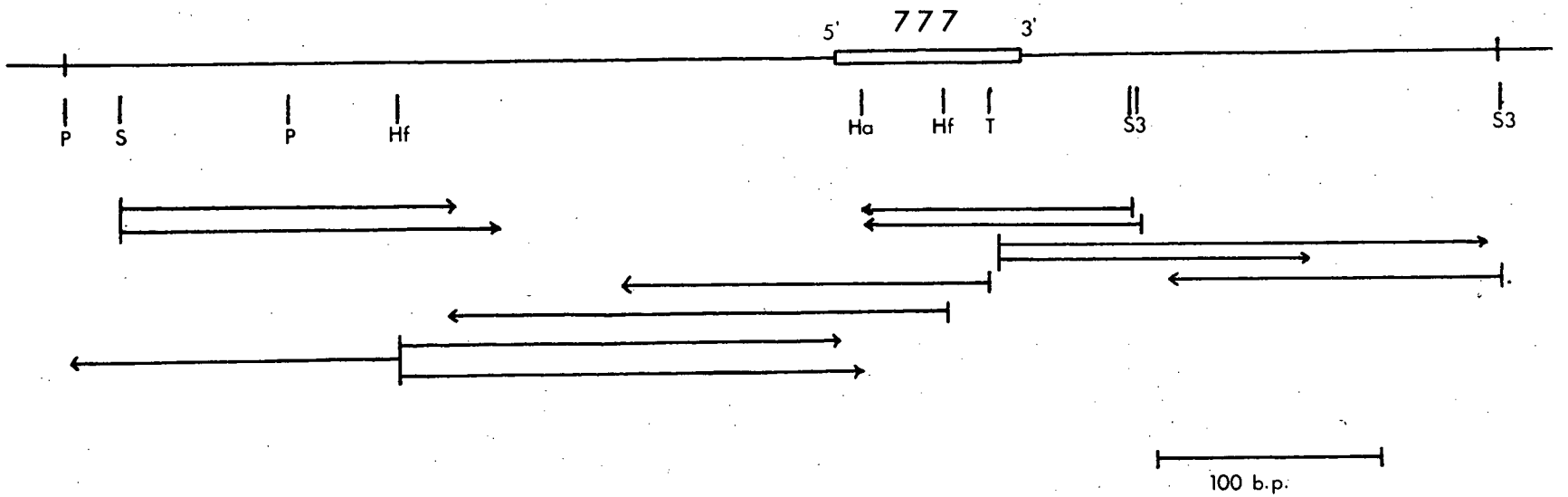
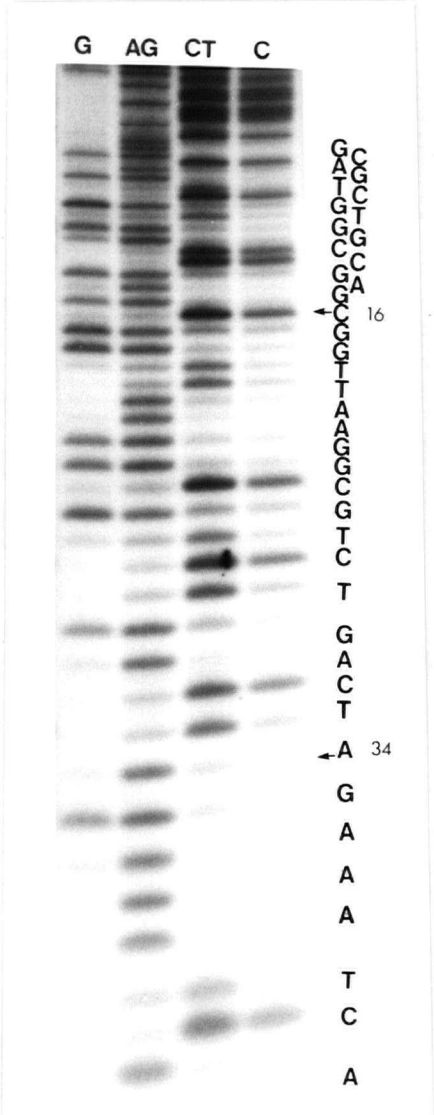


Figure 24 - Maxam/Gilbert sequence analysis of the serine tRNA gene in pDt 17R.

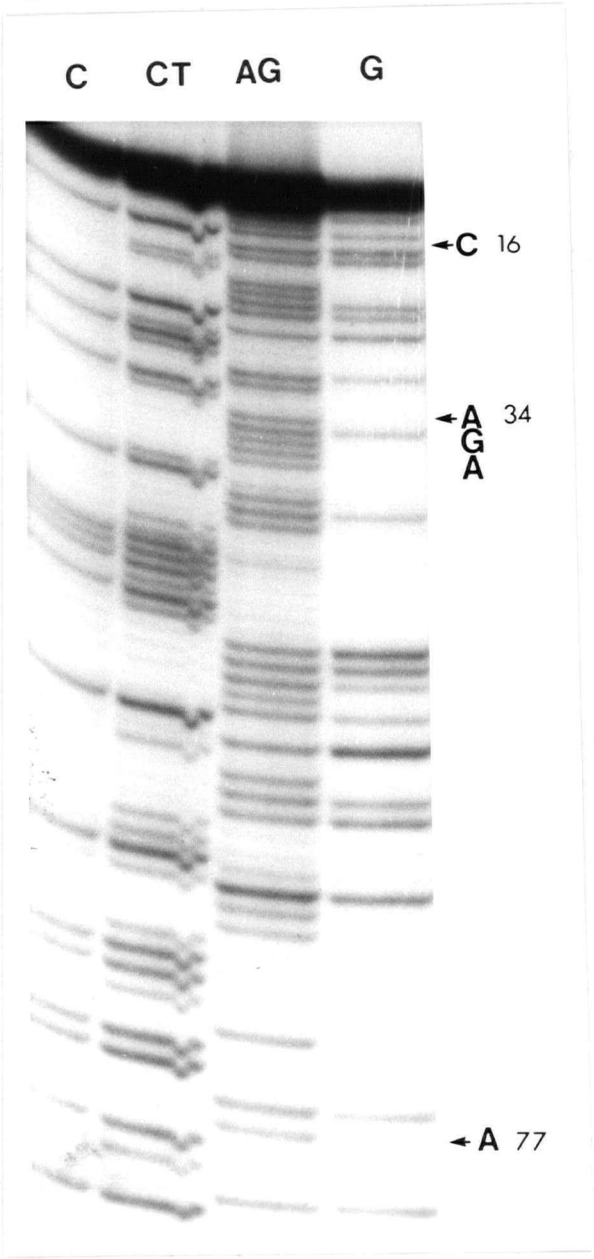
(a) Sequence analysis of a Hinf I restriction fragment labelled specifically at one end with [α-^{32}P]dTTP and E. coli DNA polymerase I (large fragment). The sequence shown is on the non-coding strand. Positions marked with arrows distinguish $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ (Figure 8).

(b) Sequence analysis of a Sau3AI/Hae III restriction fragment labelled at the Sau3AI and with [α-^{32}P]dGTP and E. coli DNA polymerase I (large fragment). The sequence shown, on the non-coding strand, includes the three positions distinguishing $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ (marked with arrows; see Figure 8).

a.



b.



fragment generated by Msp I cleavage (Fig. 16a). This estimated size is only slightly larger than that of a $tRNA^{Ser}$ coding sequence. Previous sequence analysis of a Taq I-generated fragment from pDt 16 had identified a sequence homologous to the 5' three-quarters of $tRNA_7^{Ser}$ with an Msp I site located 30 nucleotides upstream of the gene (D. Taylor, unpublished observations). The small $tRNA^{Ser}$-hybridizing Msp I fragment seen in Southern blotting experiments might then be due to a second sort of "hybrid" $tRNA^{Ser}$ gene sequence (a 774 gene). To test this possibility, pDt 16 and pBR 322 DNAs were cut with Msp I, radiolabelled with [α-^{32}P]dCTP by E. coli DNA polymerase I, and the products separated by polyacrylamide gel electrophoresis. A fragment of the predicted size (about 105 b.p.) present in pDt 16 but not pBR 322 was recovered from the gel for sequence analysis. After cleavage with Hae III (which cuts in $tRNA^{Ser}$ genes) and purification on a polyacrylamide gel, the resulting Msp I (^{32}P-dCMP)/Hae III fragments were sequenced. Positions 16 and 34 of the $tRNA^{Ser}$ sequence present in the 105 b.p. Msp I fragment were identified directly while position 77 was identified by the Msp I restriction site. In this way, one of the two $tRNA^{Ser}$ genes in pDt 16 was shown to be a "hybrid" 774 gene (Fig. 25a). In order to verify the nature of the 774 gene, and to identify the second $tRNA^{Ser}$ gene present on the plasmid, pDt 16 was characterized further.

The restriction endonuclease cleavage map of pDt 16 was determined by gel electrophoretic analysis of DNA fragments produced by digestion with one or two hexanucleotide-recognizing enzymes (Fig. 26a), and by Southern blot analysis (Fig. 26b). The restriction enzyme map for the

Figure 25 - Maxam/Gilbert sequence analysis of the serine tRNA genes in pDt 16.

(a) Sequence analysis of an Msp I/Hae III restriction fragement labelled at the Msp I end with [α-^{32}P]dCTP and E. coli DNA polymerase I (large fragment). The sequence shown is on the non-coding strand. Two of the three positions distinguishing $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ (Figure 8) are indicated by arrows. The third is identified by the presence of the labelled Msp I end. This sequence is for the 774 gene in pDt 16.

(b) Sequence analysis of a Taq I/Hae III restriction fragment labelled at the Taq end with [α-^{32}P]dCTP and E. coli DNA polymerase I (large fragment). The sequence shown is on the non-coding strand. Positions distinguishing $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ are indicated with arrows. The sequence shown is in the 777 gene of pDt 16.

(c) Sequence analysis of a Taq I/Fnu4HI restriction fragement labelled at the Taq I end with [α-^{32}P]dCTP and E. coli DNA polymerase I (large fragment). The sequence shown, on the coding strand, is complementary to $tRNA_7^{Ser}$. The arrow indicates position 77, the third position distinguishing the 777 gene of pDt 16.

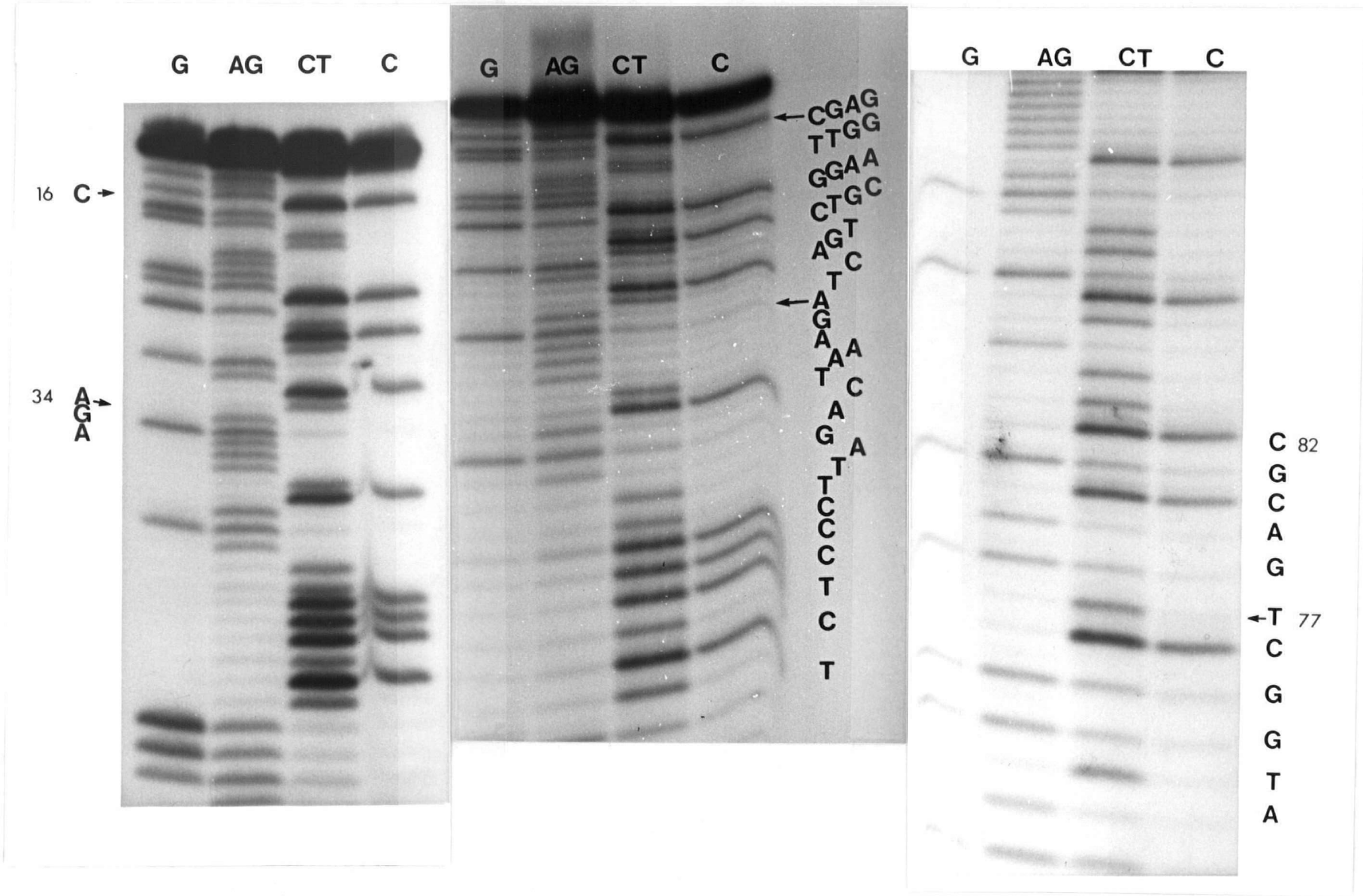
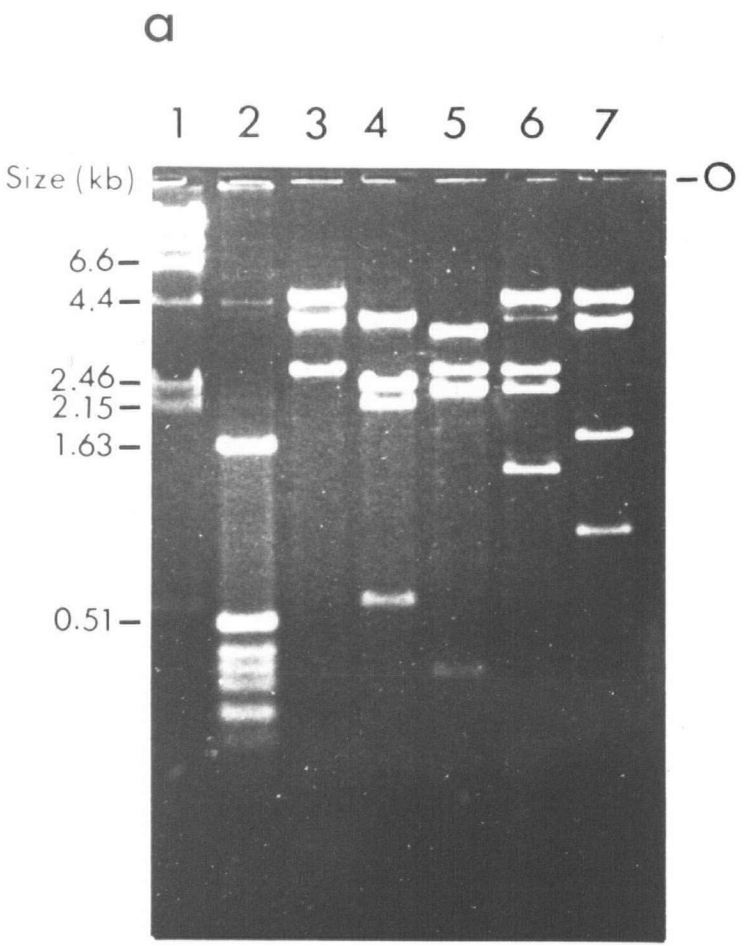
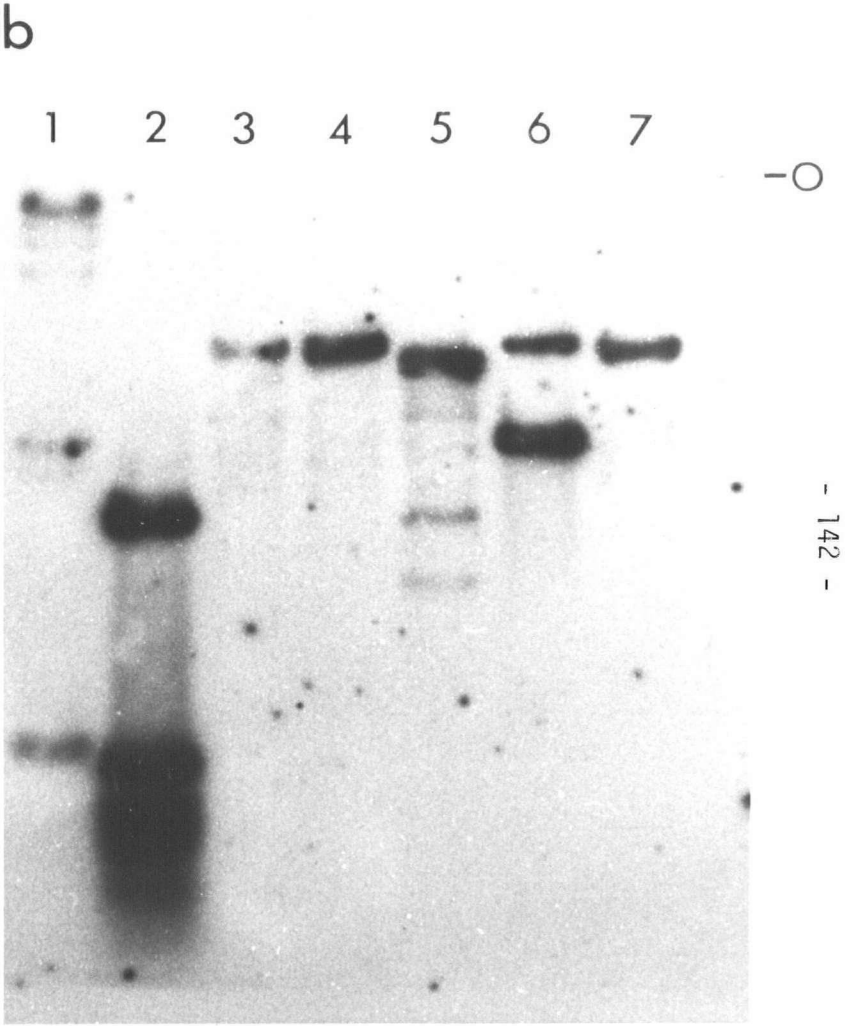

a.
b.
c.
G
AG
CT
C
16 C
34 A G A
C G A G
T T G G
G G A A
C T G C
A G T
A C
T
A
G A A A
T A C
G A
G T
T C C C C T C T
C 82
G
C
A
G
T 77
C
G
G
T
A

Figure 26 - Restriction mapping of pDt 16.

(a) DNA was digested with Pvu II (slot 3), Pvu II and Hind III (4), Pvu II and EcoR I (5), Pvu II and Bgl II (6), or Pvu II and BamHI (7), and the resulting fragments separated by electrophoresis on a 1.25% agarose gel (as described in Chapter II). Size standards were Hind III-cut bacteriophage λ DNA (slot 1) and Hinf I-cut pBR 322 DNA (slot 2) labelled with [α-^{32}P]dATP and E. coli polymerase I (large fragment).

(b) Restriction fragments from the gel shown in part (a) were transferred to nitrocellulose sheets and hybridized with [3'-^{32}P]tRNA$_7^{Ser}$ (Southern blotting, Chapter II). Autoradiography was for 2 days.


a
1 2 3 4 5 6 7
Size (kb)
6.6
4.4
2.46
2.15
1.63
0.51
O


b
1 2 3 4 5 6 7
O

insert of pDt 16 is presented in Fig. 27. The two $tRNA^{Ser}$-hybridizing sequences in pDt 16 (Section A.1. above) are located between the EcoR I and Bgl II sites in the insert.

The 3.6 kb Pvu II fragment containing the $tRNA^{Ser}$ genes was purified by agarose gel electrophoresis for use in mapping and sequence analysis. Cleavage of this fragment with Bgl II, radiolabelling of the Bgl II ends, and separation of the resulting fragments by gel electrophoresis gave two ^{32}P-labelled fragments of roughly the expected sizes (2.3 kb and 1.3 kb) plus a third small (~ 150 b.p.) fragment. This indicates the presence of two closely spaced Bgl II sites in the insert, similar to the two closely-spaced Pvu II sites near the $tRNA^{Ser}$ gene in pDt 17R. In neither case was the second restriction site observed in preliminary mapping procedures.

The 2.3 kb Bgl II (^{32}P-dGMP)/Pvu II fragment, which was expected to contain the $tRNA^{Ser}$ genes, was mapped further by the method of Smith and Birnstiel (151). Ten different enzymes were used in two mapping experiments. The locations and orientations of the two $tRNA^{Ser}$ genes could be discerned, as well as a number of surrounding restriction sites (see figure 28). The two genes are present as a direct repeat separated by a spacer region of about 400 b.p. The 5' ends of the genes are located about 0.5 kb and 1.0 kb from the labelled Bgl II site. The former gene was deduced to be either 477 or 777 by the two Hinf I-generated bands, one Taq I band, and absence of a band in the Msp I and Fnu4H I slots (Fig. 28). The latter gene is the 774 sequence. It was distinguished as a region containing two closely spaced Msp I sites, plus two Hinf I sites, a Taq I site, a Fnu4H I site, and an Alu I site (known to be present from the unpublished results of D. Taylor and from sequence analysis of

Figure 27 - Restriction enzyme map of the Drosophila insert in pDt 16.

Abbreviations used for restriction enzymes are E (EcoRI), Hd (Hind III), P (Pvu II), B (BamHI), and Bg (Bgl II). Heavy lines at each end represent pBR 322 sequences.

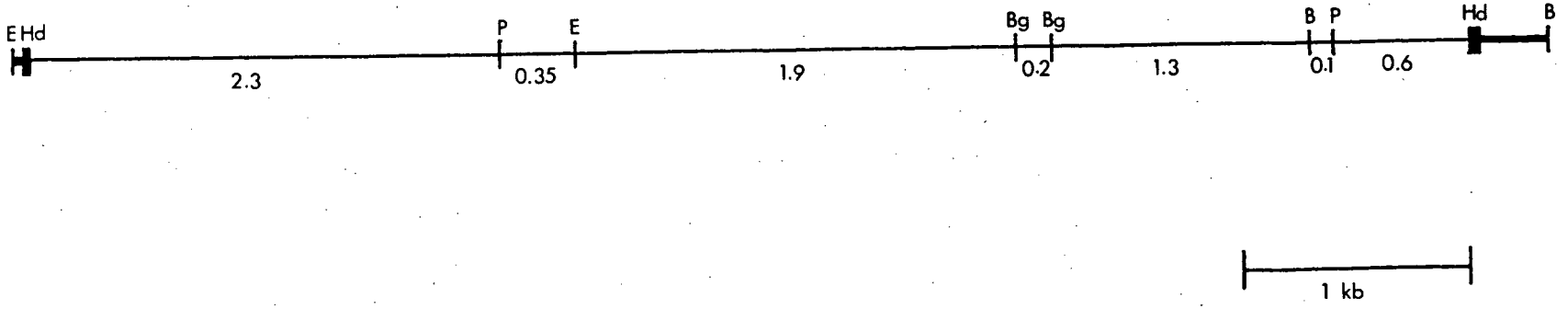
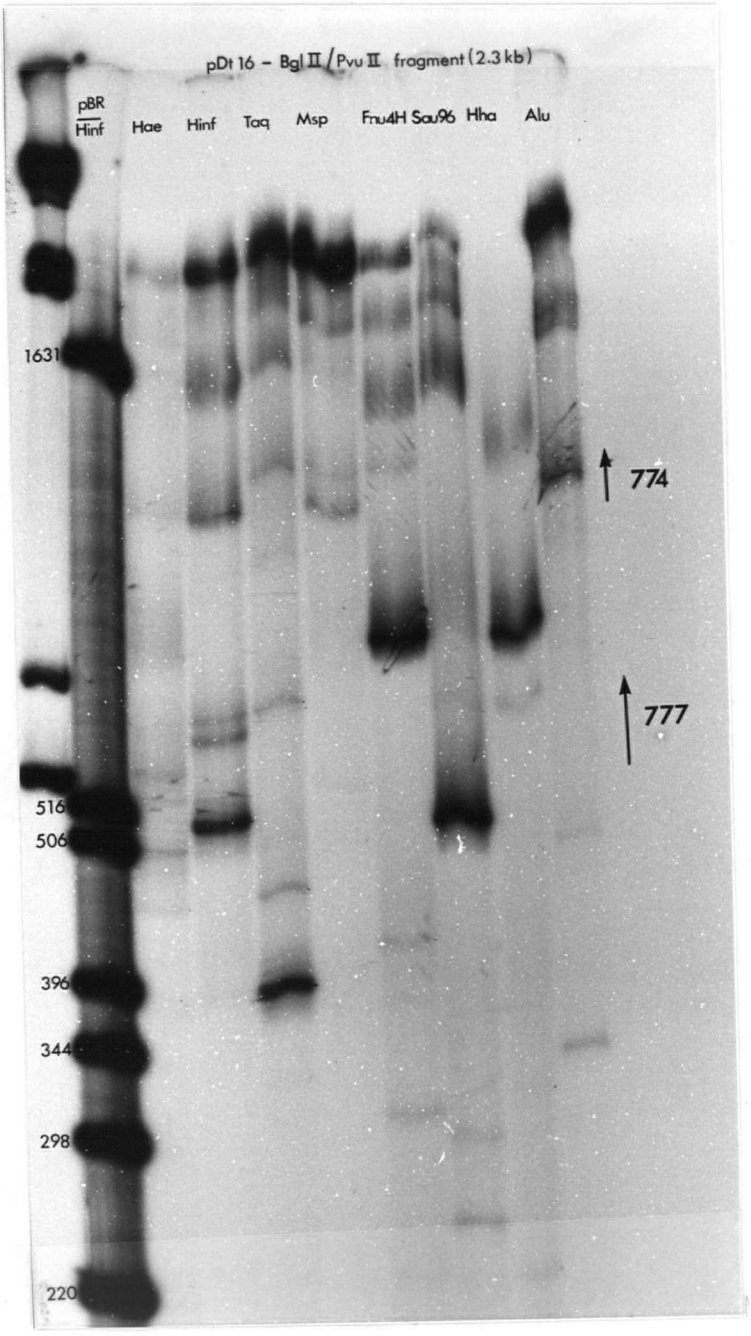

E Hd
P
E
Bg Bg
B P
Hd
B
2.3
0.35
1.9
0.2
1.3
0.1
0.6
1 kb

Figure 28 - Smith/Birnstiel restriction mapping of the gene-containing region in pDt 16.

Mapping of a 2.3 kb Bgl II/Pvu II restriction fragment labelled at the Bgl II site with [α-^{32}P]dGTP and E. coli DNA polymerase I was accomplished using the method of Smith and Birnstiel (Chapter II). The fragment was partially digested with Hae III (Hae), Hinf I (Hinf), Taq I (Taq), Msp I (Msp), Fnu4H I (Fnu 4H), Sau96 I (Sau96), Hha I (Hha), or Alu I (Alu). The size standards were Hinf I-derived fragments of pBR 322 labelled with [α-^{32}P]dATP and E. coli DNA polymerase I (pBR/Hinf). Electrophoresis was for 16 hr. at 125V on a 3% polyacrylamide gel. Each gel slot contained 42,000 Cerenkov cpm. Autoradiography was for 39 hr. at 4° using Gevaert "screen" film.


pDt 16 – Bgl II / Pvu II fragment (2.3 kb)
pBR Hinf
Hae
Hinf
Taq
Msp
Fnu4H
Sau96
Hha
Alu
1631
516
506
396
344
298
220
774
777

the 105 b.p. Msp I fragment, above). Little useful map information could be deduced further from the Bgl II site than the 774 gene.

The 3.6 kb Pvu II fragment was digested with Taq I, and the fragments labelled with [α-^{32}P]dCTP with the Klenow fragment of E. coli DNA polymerase I. Portions of the radiolabelled DNA were digested with a number of other restriction enzymes and the products separated by polyacrylamide gel electrophoresis. In conjunction with Smith/Birnstiel mapping (above) and predicted restriction maps based on $tRNA^{Ser}_{4,7}$ sequences (Fig. 17), the gene containing fragments could be identified.

Mapping of the 3.6 kb Pvu II fragment was also attempted by digesting the DNA with Fnu4H I ($GC^{\downarrow}NGC$) and 3'-end labelling with the Klenow fragment of E. coli DNA polymerase I and (α-^{32}P) labelled dATP, dGTP, dCTP, or dTTP. The one nucleotide 5'-extension produced by cleavage with this enzyme allows specific labelling of every fragment end with only one of the four [^{32}P] deoxynucleotides. In some cases the ordering of fragments can be deduced, since the two fragment ends produced by a single cleavage with Fnu4H I must be labelled by complementary nucleotides (dATP and dTTP, or dGTP and dCTP). Also, such mapping allows one to discern which fragments are single-end labelled (and therefore ready for sequence analysis) or double-end labelled by a particular [α-^{32}P] dNTP and E. coli DNA polymerase I.

DNA sequence analysis of fragments produced by Taq I (Fig. 25b,c) or Fnu4H I digestion of the 3.6 kb Pvu II fragment allowed a nearly continuous DNA sequence of 530 b.p. to be determined, extending from the Hae III site at the 5'-end of the 777 gene to the 3' end of the 774 gene

gene (Fig. 29a). Most of the sequence was obtained only once. In the places where a sequence was determined twice, on one or both strands, good agreement was obtained. The sequence of the first gene, though not complete, includes the three positions distinguishing $tRNA_7^{Ser}$ and $tRNA_4^{Ser}$, and especially position 16, which is not part of a restriction enzyme recognition sequence (Fig. 17). The sequence determined corresponds to a 777 gene (Fig. 25 b,c), in agreement with the results of Smith/Birnstiel mapping. The CCA of the mature tRNA is not gene-coded. All except the 3'-terminal nucleotide of the 774 gene either were determined directly by DNA sequence analysis, or could be deduced on the basis of Msp I and Fnu4H I specificities. The three positions distinguishing $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ genes were all determined at least twice (the 3'-end by cleavage with both Msp I and Fnu4H I). The extents and directions of sequences obtained in Maxam-Gilbert sequence experiments are shown in Fig. 29b. The nucleotide sequence of this region is compiled in Fig. 29a.

A.6. Summary of results from DNA sequence analysis of pDt 16, pDt 17R, and pDt 73

Determination of the complete nucleotide sequences of two genes (pDt 73 and pDt 17R), and mapping and partial nucleotide sequence determination for two other genes (pDt 16), showed the presence of three different sequences related to $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ in a distinctive way. The single gene in pDt 17R corresponds exactly to $tRNA_7^{Ser}$ (777 gene). The single 474 gene in pDt 73, however, is intermediate between the sequences expected in $tRNA_4^{Ser}$ (444) and $tRNA_7^{Ser}$ (777) genes. Restriction mapping and DNA sequence analysis allowed identification of the two $tRNA^{Ser}$ sequences in

Figure 29 - Nucleotide sequence of the tRNA gene-containing region in pDt 16.

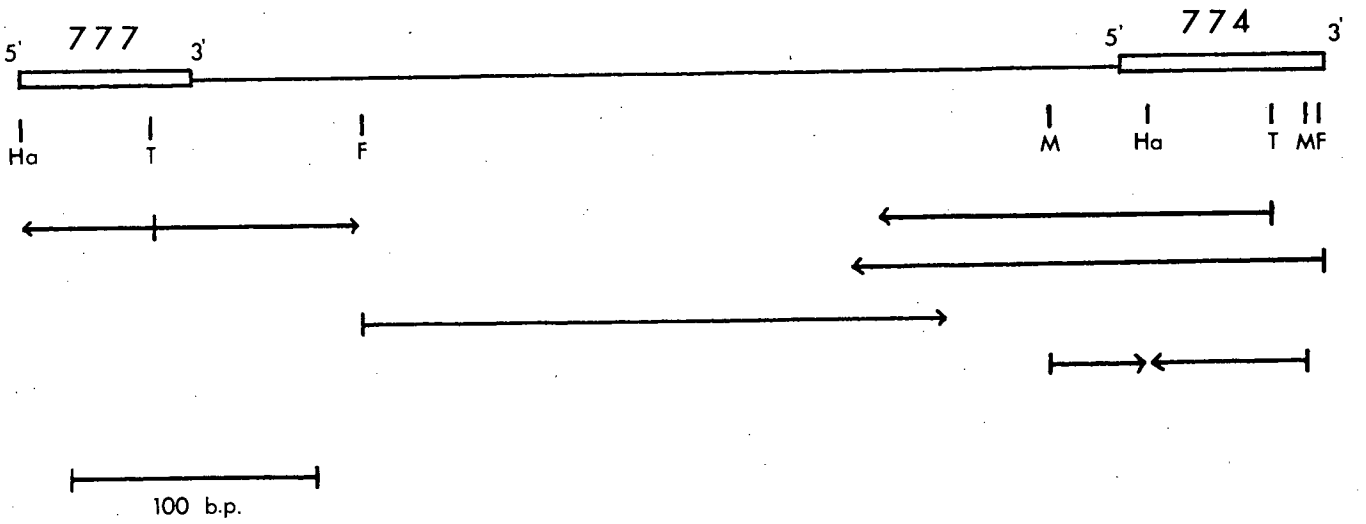
(a) The compiled sequence obtained from the gene-containing region of pDt 16 is shown (non-coding strand for both $tRNA^{Ser}$ genes). The 777 and 774 gene sequences are underlined.

(b) A restriction map of the gene-containing region in pDt 16, indicating the extents and directions of sequences obtained in Maxam/Gilbert sequencing experiments, is shown. A number of relevant restriction enzyme cleavage sites are indicated: F (Fnu4H I), Ha (Hae III), M (Msp I), and T (Taq I).

a. The Nucleotide Sequence of pDt16

```
        10         20         30         40         50         60         70         80         90        100
GGCCGAGCGG TTAAGGCGTC TGACTAGAAA TCAGATTCCC TCTGGGAGCG TAGGTTCGAA TCCTACCGAC TGCGAATAGC AATCTGTTTT TTGGAAGTCC

       110        120        130        140        150        160        170        180        190        200
AGAAAAAATA GATCGCTAGA AGATCAGAAA AAGTATTAAG AGCGCTGCTC TCTTATAATG CTTAAAAAAT ATTTCGTAGT AAAAGAGTAA AGTGTGTGGC

       210        220        230        240        250        260        270        280        290        300
AAATAAAATC ATGCACCTTT GTAAAGTTAC TGATATAAAG AAAGAATTTA ATGATTTAAA AAATAGCCCA ACAAGTTGGG TGATCTCTTA AAAATACAGT

       310        320        330        340        350        360        370        380        390        400
GCTGAAACCA ATTTAACTTT TTTGAATTTA ATCATTATCT ATTGTTAAAA AAGTGATATT AATAGTTATA TGATCGTCTT TTCGCTATAA AAAGATCAGT

       410        420        430        440        450        460        470        480        490        500
GATATTAATG TAGCTAGAGC CGGGTAATAA AGCCACTAGA GTCATCAAAG CAGTCGTGGC CGAGCGGTTA AGGCGTCTGA CTAGAAATCA GATTCCCTCT

       510        520        530
GGGAGCGTAG GTTCGAATCC TACCGGCTGC
```

b.



pDt 16. This plasmid contains both a 777 and a 774 gene as a direct repeat separated by a 376 b.p. spacer region. Thus, in the four genes analyzed, three different but closely related sequences were found, of which only one corresponds to a known $tRNA^{Ser}$ sequence. The existence of a fourth closely related $tRNA^{Ser}$ sequence is presumed, based on the known sequence of $tRNA_4^{Ser}$.

A.7. Comparison of gene-flanking sequences

A.7.a. 3'-flanking sequences

Extended sequences adjacent to the 3' ends of $tRNA^{Ser}$ genes were obtained only for the 777 genes in pDt 16 and pDt 17R. No sequence 3' to the 774 gene of pDt 16 was obtained, while the presence of the Hind III site defining the end of the pDt 73 insert only five nucleotides from the 474 gene allowed only eleven nucleotides of flanking sequence to be determined. This flanking sequence does not extend to a series of T residues (non-coding, or "+" strand) thought to signal RNA polymerase III transcription termination. Therefore this flanking sequence is not discussed further here.

The sequences flanking the 3' ends of 777 genes in pDt 17R and pDt 16 share limited homology when directly compared (see Figure 30a). Both sequences contain tracts of six consecutive T residues in the non-coding strand beginning 13 and 15 nucleotides after the genes in pDt 16 and pDt 17R, respectively. These are presumably RNA polymerase III termination signals (163). Both also contain five nucleotide homologies

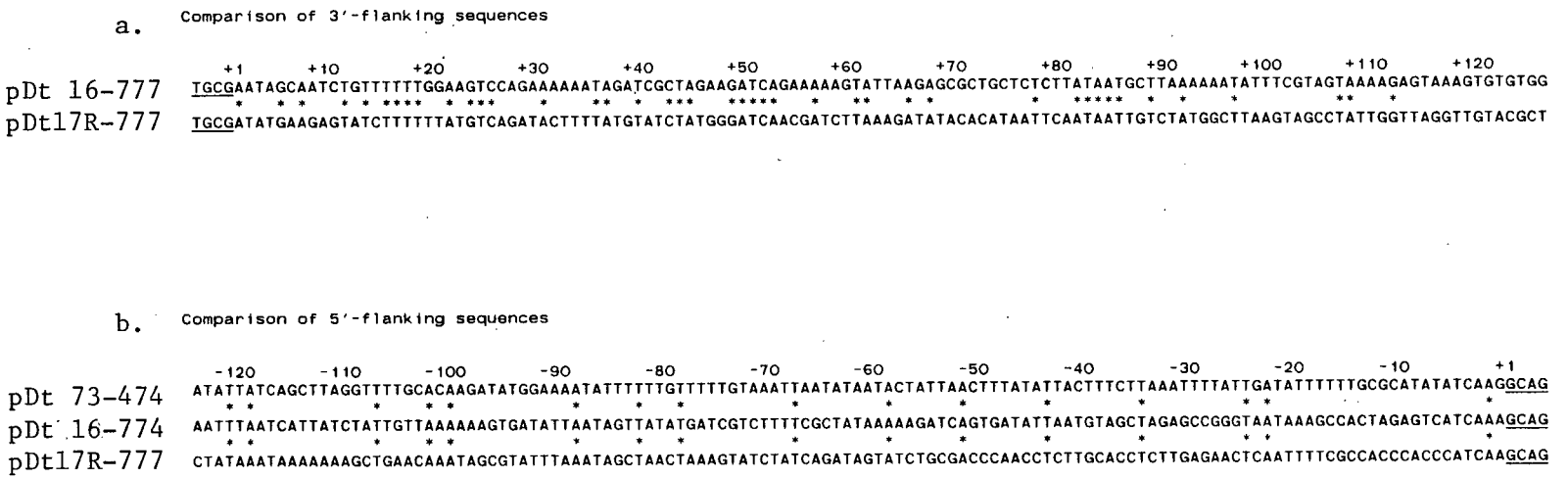


Figure 30- Comparison of serine tRNA gene-flanking sequences.

at 48-52 and 81-85 nucleotides after the genes, and four of five at 21-25 nucleotides after the genes. The first 125 nucleotides after the two 777 genes share 33% homology (42 of 125). Most of this homology is apparently a consequence of the general A-T richness of the DNA sequences: only six of 42 homologous positions in the regions compared contain G-C base pairs, and four of the six are located in the short homologies noted above. The flanking regions examined contain only 31% (pDt 16) and 30% (pDt 17R) G-C base pairs. Determination of more $tRNA^{Ser}$-flanking sequences might clarify whether the small clusters of homology observed are significant.

A.7.b. 5'-flanking sequences

The first 125 nucleotides of 5'-flanking sequences for 777 (pDt 17R), 774 (pDt 16), and 474 (pDt 73) genes were examined for homology (Fig. 30b). The sequences contain a low fraction of G-C base pairs (pDt 16 contains 28% G-C; pDt 17R, 37%; and pDt 73, a very low 19%). Little similarity was observed among the three sequences. By comparison of sequences aligned with respect to the $tRNA^{Ser}$ genes (Fig. 30b), 13 of 125 positions were found to contain the same nucleotide in all three sequences. In all cases, the position contained on A-T base pair. In no case were two consecutive nucleotides maintained among the three sequences when aligned in this fashion.

Certain features located near the sequence coding for mature $tRNA^{Ser}$ were of interest. All three sequences contained the pentamer ATCAA immediately adjacent to the tRNA sequence or one nucleotide removed from it (Fig. 30b). The pentamer GCCAC appears at -14 to -18 relative to

the 774 gene in pDt 16, and at -12 to -16 relative to the 777 gene in pDt 17R. A similarly G-C rich, through not homologous, sequence (GCGCA) is found at -10 to -14 relative to the 474 gene in pDt 73. These conserved sequences may serve a specific function, for example in transcription initiation or processing of a tRNA precursor. However, assignment of such a function awaits direct experimental evidence.

B. Discussion

The DNA sequences of four $tRNA^{Ser}$ genes from Drosophila melanogaster were analyzed. The major result of this work was the finding that several different closely related $tRNA^{Ser}$ sequences are present in the D. melanogaster genome. Genes corresponding to mature $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ are designated 444 and 777 genes, respectively, on the basis of the three differing positions in the two tRNA sequences. In the analyses presented above, two 777 sequences were found as well as single 774 and 474 sequences. Since $tRNA_4^{Ser}$ exists, a corresponding 444 gene must be present in the genome. Thus, at least four different but closely related $tRNA^{Ser}$ gene sequences are represented in the D. melanogaster genome.

Is sequence heterogeneity normal for reiterated eukaryotic tRNA genes?

Based largely on the results of studies on tRNA-DNA hybridization kinetics, it seems to have been widely expected that members of reiterated gene families would be identical. This expectation has largely been supported by the results of DNA sequence analysis of tRNA genes, as

reviewed in Chapter I. However, a number of examples of closely related but non-identical tRNA genes have been reported. Included in this category are D. melanogaster gene sequences closely related but not identical to $tRNA_i^{Met}$ (121), $tRNA_4^{Val}$ (122), $tRNA^{Glu}$ (119), and $tRNA_5^{Lys}$ (124). The heterogeneity seen for $tRNA^{Ser}$ genes in the present work, though relatively extreme, is consistent with the results of other workers mentioned above. Because $tRNA^{Ser}$ genes from the 12DE region of the D. melanogaster chromosome represent the most extreme exception observed to date to the generalization that reiterated tRNA genes are identical, they may prove to be a useful model system for studying whether and how the identity of redundant genes is maintained.

Are the "hybrid" serine tRNA genes expressed?

It is not yet known whether the 774 and 474 serine tRNA genes are transcribed. However, several observations suggest that they are likely to be. A number of examples of closely related isoacceptor tRNA sequences (≥ 95% homologous) have been reported. These are presented in Fig. 31. Such tRNAs must arise from transcription of closely related genes. (Also, their existence suggests that tRNA gene heterogeneity is not uncommon, relevant to the previous paragraph.) It is possible that the hybrid $tRNA^{Ser}$ genes may correspond to $tRNA_6^{Ser}$, $tRNA_1^{Ser}$, or $tRNA_3^{Ser}$. Particularly likely candidates would seem to be $tRNA_6^{Ser}$ and $tRNA_3^{Ser}$, as their chromatographic properties on RPC-5 columns are quite similar to those of $tRNA_7^{Ser}$ and $tRNA_4^{Ser}$, respectively. The presence of the conserved pentameric sequence ATCAA preceding 777, 774, and 474 genes suggests that this sequence may

Figure 31 - Closely related but non-identical tRNAs.

Same anticodon	$tRNA^{Tyr}_{1,2}$ *E. coli* (2 differences)	[a]
	$tRNA^{Arg}_{3a,3b}$ of yeast (2 ")	[a]
	$tRNA^{Thr}_{1a,1b}$ of yeast (2 ")	[a]
	$tRNA^{Ser}_{1,2}$ of yeast (3 ")	[a]
	$tRNA^{Ala}_{1,2}$ of yeast (1 ")	[a]
	$tRNAs^{Leu}$ of hepatoma (1 ")	[a]
	$tRNA^{Lys}_{1,2}$ of rabbit (2 ")	[a]
	$tRNA^{Phe}_{1,2}$ of bovine lens (1 ")	[a]
Different anticodons	$tRNA^{Ser}_{minor}$ of yeast (3 ")	[b]
	$tRNA^{Ser}_{4,7}$ of *Drosophila* (3 ")	[c]
	$tRNA^{Ser}_{2a,1}$ of rat liver (3 ")	[d]

a - Reference 7.

b - Reference 67.

c - Chapter III.

d - Reference 176.

be involved in transcription initiation or in processing of precursor $tRNA^{Ser}$. It seems unlikely that transcripts from 774 or 474 genes would be preferentially degraded simply because the mature gene products would be so similar to $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$. (Preferential degradation of precursors from these genes cannot be ruled out.) Further, it seems unlikely that the presence of the correct nucleotides at both positions 16 and 77 is required for functional split promoters within 474 or 774 genes (70). In any case, such an argument would only apply to 774 genes, since with respect to the relevant nucleotides located in putative promoter regions a 474 gene is equivalent to a 444 gene, which is expressed as $tRNA_4^{Ser}$. The question of whether "hybrid" $tRNA^{Ser}$ genes are transcribed could be tested by transcription of pDt 16 (which contains both a 777 and a 774 gene) in vitro in recently developed homologous Drosophila systems (129,164), or by sequence analysis of $tRNA_6^{Ser}$, $tRNA_1^{Ser}$, and/or $tRNA_3^{Ser}$.

What is the significance of "hybrid" $tRNA^{Ser}$ genes in Drosophila?

A mechanism presumably exists for maintaining families of reiterated tRNA genes in a homogeneous state. This statement is based on the numerous examples of identical tRNA genes presented in sections C and D of Chapter I. It seems likely that such a mechanism should match pairs of genes through recombination. Such recombination should be non-reciprocal: a reciprocal process would not normally result in maintenance of a wild-type sequence. The dispersed organization of tRNA genes within a cluster argues against unequal crossover as a mechanism for maintaining these genes in D. melanogaster.

The distinctive relationship existing among $tRNA^{Ser}$ genes in D. melanogaster suggests that 777, 774, 474, and 444 genes are products of the action of such a maintenance process. It is true that the "hybrid" 774 and 474 sequences can be simply explained as the products of reciprocal recombination events involving the genes for $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$. However, the high degree of similarity between $tRNAs^{Ser}$ with anticodons CGA and IGA in D. melanogaster and in rat liver, and between the minor $tRNAs^{Ser}$ with anticodons CGA and U*GA in S. cerevisiae is consistent with the idea that the corresponding sets of genes have not been free to diverge during evolution (see also Discussion, Chapter III). Reciprocal recombination seems to have no bearing on this observation. On the other hand, if very similar but non-identical genes with different anticodons were regularly checked against each other in the maintenance process, they would be expected to evolve in parallel. It is clear that such closely related $tRNA^{Ser}$ sequences are not necessary for proper function in protein biosynthesis, since $tRNAs^{Ser}$ responding to UCN and AGPy codons have relatively dissimilar sequences (Chapter III). Further, none of the four $tRNA^{Ser}$ genes examined here contains a nucleotide in it which is not distinctive of either $tRNA_4^{Ser}$ or $tRNA_7^{Ser}$. It seems quite possible that the very high degree of similarity among the three types of serine tRNA genes found in plasmids derived from the 12DE region and $tRNA_4^{Ser}$ genes may result in aberrant maintenance of the genes, and in the "mingling" of $tRNA_{4,7}^{Ser}$ sequences in the genes. (This last statement implies that a maintenance process operates for tRNA genes on the X chromosome as well as on the autosomes.) Still another possibility is that the "hybrid" genes observed are the products of recent reciprocal crossover events in opposition to a maintenance process.

The serine tRNA gene sequences in the 12DE region may or may not be stably maintained over long time periods. In view of the similarity of the 777, 774, and 474 sequences, it seems doubtful to me that they could exist stably over evolutionary time periods in the presence of the presumptive mechanism responsible for maintaining various other reiterated genes in an identical state. Whether individual tRNA genes are static or dynamically changing due to the presumptive maintenance process could be testing by probing genomic DNA for the presence or the absence of an Msp I site in individual $tRNA^{Ser}$ genes in Southern blotting experiments. The probes would be unique sequences flanking the 5' or 3' ends of $tRNA^{Ser}$ genes in recombinant plasmids analyzed here. In principle, such a procedure could be used to test genomic DNA isolated from individual flies or from populations. If the maintenance process is dynamically interconverting $tRNA^{Ser}$ genes, it might prove possible by such methods to identify two different Msp I fragments in genomic D. melanogaster DNA which both hybridize the specific probe and differ only by the presence or absence of an Msp I restriction site within a $tRNA^{Ser}$ gene. Further, by using both 5' and 3' gene-flanking probes such experiments should allow one to distinguish between non-reciprocal and reciprocal recombination events involving those $tRNA^{Ser}$ genes (within the limitation that reciprocal double-crossover events occurring in a very short sequence would appear non-reciprocal).

A convincing biochemical demonstration that maintenance of tRNA gene families is a dynamic process might be achieved by isolating two distinct forms of the same gene independently in recombinant plasmids that are otherwise identical. This sort of demonstration could possibly

be achieved experimentally with S. cerevisiae by integrating a recombinant plasmid carrying a wild-type tRNA gene into the yeast genome adjacent to a suppressor tRNA gene, then re-isolating the same recombinant plasmid from the yeast (165). By selecting for yeast strains with enhanced suppressor function, and re-isolating the integrated plasmid from the yeast genome, it might be possible to demonstrate directly a specific alteration of a wild-type tRNA gene to a suppressor tRNA gene.

The structures of the $tRNA^{Ser}$ genes analyzed above suggest that these genes are the result of a dynamic maintenance process. This remains to be shown experimentally. Based on the limited number of genes examined, the existence of 747 and 447 genes produced in the same events as 474 and 774 by reciprocal recombination between 444 and 777 genes cannot be discounted. Such diversity cannot be readily explained without assuming that one of the 747 or 447 genes is inactive. Otherwise, reciprocal recombination would result in six genes producing $tRNAs^{Ser}$ reading UCN codons. There do not appear to be enough $tRNA^{Ser}_{UCN}$ species in D. melanogaster to accommodate such diversity. No more than seven serine-accepting tRNAs are distinguished on RPC-5 columns, of which two respond to AGPy codons (102). Further, such a reciprocal process would not limit the evolutionary divergence of two closely related tRNA species, as discussed above. In the following discussion, I assume that the similarity of the 444 and 777 genes, and the existence of "hybrid" 474 and 774 genes, is a consequence of the operation of a non-reciprocal maintenance process; and that these sequences are at least in part interchangeable in the process.

Gene maintenance might occur only within a single gene cluster, or also between genes at different chromosomal locations. It seems likely

that the frequency of matching between two dispersed genes would be considerably less than that between genes within a single cluster. Thus, the presence of nucleotides distinctive of $tRNA_4^{Ser}$ in two of the four $tRNA^{Ser}$ genes examined (77$\underline{4}$ and $\underline{4}$7$\underline{4}$) may indicate that at least one 444 gene is present in the 12DE gene cluster. A dynamic maintenance process would probably be bi-directional (unless some mechanism exists for distinguishing "old" correct DNA sequences from "new" incorrect ones.) Formation of either of two possible pairs of identical products ("products" refers to the scope of an individual maintenance event, which might span only a few or many nucleotides) would be equally likely. Thus, maintenance of a sequence present in multiple copies would be favored over one present in a single copy. As each of the four $tRNA^{Ser}$ genes examined contains the same anticodon, this last point may not be significant here in terms of biological function of tRNA. However, such a consideration may be important in terms of genes containing a CGA anticodon (e.g. 444 genes) rather than AGA present in 777, 774, and 474 genes. Reduction in the number of genes with a particular anticodon such as CGA might result in decreased viability of the organism. (The same applies to genes with AGA anticodons, of course. However, there does not seem to be any shortage of this anticodon based on the limited number of genes sequenced.) Alternatively, some anticodons such as CGA could conceivably be protected in some fashion against frequent maintenance. Along these lines, it is interesting to consider the genes in S. cerevisiae coding for minor serine tRNAs that read UCA and UCG codons. Three identical gene copies for $tRNA_{UCA}^{Ser}$ and a single gene copy for $tRNA_{UCG}^{Ser}$ are present in the haploid genome of this yeast (67). Mutations in the anticodon of the UCG-specific tRNA gene are lethal in the haploid cell, and must be maintained in diploids (67,166).

The two mature $tRNAs^{Ser}$ encoded by these genes differ at only three positions. One is the first position of the anticodon. The other two result from a G-C to C-G transversion of one base pair in the anticodon stem. Thus, they are as similar to each other as $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ of D. melanogaster are. The single gene for $tRNA_{UCG}^{Ser}$ contains an intervening sequence adjacent to the anticodon, whereas the three genes for $tRNA_{UCA}^{Ser}$ do not. One might speculate that one possible function for the intervening sequence in the $tRNA_{UCG}^{Ser}$ gene is to protect the anticodon against change by preventing matching of UCA and UCG-specific anticodon sequences. The recognizable homology of intervening sequences in $tRNA^{Leu}$ genes from S. cerevisiae and D. melanogaster (87) is consistent with conservation of a functional DNA sequence. Protection of the anticodon sequence may be one such function. Such speculative arguments would be strengthened if equivalent $tRNA_{UCG}^{Ser}$ genes in Drosophila and rat liver are found to contain intervening sequences.

Gene conversion as a possible mechanism for maintaining reiterated genes

By analogy with yeast, a possible mechanism for maintaining identical reiterated tRNA genes in D. melanogaster is gene conversion. Integrative transformation of the yeast S. cerevisiae by recombinant plasmids carrying a yeast gene results in formation of a non-tandem gene repeat (167). Integration of a wild-type gene into the genome of a yeast strain carrying a mutant allele of the gene allows one to select for and characterize spores lacking gene function. Using such methods, Scherer and Davis (168), Klein and Petes (169), and Jackson and Fink (170) have found that non-reciprocal gene conversion events resulting in two mutant

gene copies occur frequently. Indeed, meiotic gene conversion may occur with a frequency as high as 2-4% of all tetrads isolated from such a yeast strain (169). These authors point out that such conversion could play an important role in maintaining reiterated gene copies in an identical state (168,169,170).

Ernst, Stewart, and Sherman (171) have found that a mutant CYC 1 gene in S. cerevisiae can revert by recombination with the non-allelic CYC 7 gene, resulting in a "composite" structural gene containing sequences from both CYC 1 and CYC 7. These authors suggest that the genetic exchange observed between non-allelic genes may be similar to the exchanges resulting in conversion of duplicated LEU 2 or HIS 4 genes (169,170). The "composite" CYC 1/CYC 7 genes and the process by which they are formed may well be analagous to the case of "hybrid" serine tRNA genes in D. melanogaster. Interestingly, the length of the "composite" sequence in reverted CYC 1 genes is in the range 43-71 b.p. (171), a distance range easily applied, by analogy, to $tRNA^{Ser}$ genes to explain the "hybrid" sequences observed.

Hottinger and Leupold (1972) observed frequent reversion of frameshift suppressor mutations in general (meiotic frequency ~ 10^{-5} spores) and an extremely high rate of reversion of frameshift suppressor mutations at the suf 5 locus (meiotic frequency ~ 10^{-3} spores) in the yeast S. pombe. They suggest that the suf 5 locus may contain two copies of the same tRNA gene sequence, and that reversion to the suppressor-inactive form could result from frequent heterologous recombinations between the adjacent tRNA genes. (As such recombinations would result in loss of suppressor function from either tRNA gene, they would be non-reciprocal and thus gene conversion events.)

It has been suggested that gene conversion is responsible for the high degree of similarity between a pair of duplicated human γ globin genes (173,174). While flanking sequences diverge considerably, the two fetal globin genes have been maintained in a near-identical state. Baltimore has discussed the possible involvement of gene conversion in maintaining immunoglobulin genes (175). Both Baltimore (175) and Ernst et al. (171) point out the possible importance of gene conversion, both in maintaining the homogeneity of reiterated gene families and in generating protein diversity during evolution.

To identify formally the presumptive process responsible for maintenance of reiterated genes in Drosophila as gene conversion would require that a non-reciprocal genetic event be demonstrated. This cannot be done, since one cannot be sure of obtaining all possible products of a meiotic event. However, it can be said in the present case of serine tRNA genes from the 12DE gene cluster of D. melanogaster that the data are quite consistent with possible operation of such a process.

Concluding Statement

Evidence for four closely related types of serine tRNA genes in Drosophila melanogaster has been found. The existence of these four sequences in the D. melanogaster genome may reflect the dynamic function of a process responsible for maintaining the homogeneity of reiterated genes. The heterogeneity observed in the work presented here is consistent with the results of other workers studying families of tRNA genes, though relatively rather an extreme case. It is suggested that due to the

considerable similarity of the various $tRNA^{Ser}$ gene sequences, the genes can be interconverted by this maintenance process, which may be analogous to gene conversion observed in yeast. Thus, serine tRNA genes may prove to be a useful model system for the study of the maintenance process in D. melanogaster, as well as for the study of tRNA gene expression.

Chapter V

SEQUENCE ANALYSIS OF $tRNA_5^{Lys}$ FROM DROSOPHILA

A. Introduction

Studies on organization of eukaryotic tRNA genes have been greatly aided by recently developed recombinant DNA techniques (177) and rapid DNA sequence analysis (132,178). Expression of these genes may be studied using *in vivo* or *in vitro* transcription systems (179,128,129,164). Such studies depend on knowing the nucleotide sequence not only of the DNA but also of the mature tRNA product. In *Drosophila melanogaster*, the two lysine codons (AAA, AAG) are read by two predominant lysine tRNAs, $tRNA_2^{Lys}$ and $tRNA_5^{Lys}$. The former has been sequenced (104), as have a number of its genes (116,123). Transcription of recombinant plasmids containing normal or rearranged $tRNA_2^{Lys}$ genes has also been examined (123). Genes for $tRNA_5^{Lys}$ have been isolated in recombinant plasmids (114). As part of a study on lysine tRNAs and their genes in *D. melanogaster*, the nucleotide sequence of $tRNA_5^{Lys}$ was examined.

Relevant to the subject of Chapters III and IV above, $tRNA_5^{Lys}$ hybridizes to the 12DE region of polytene chromosomes under some conditions (109). Thus, the $tRNA_5^{Lys}$ sequence determination presented in this chapter also represents a part of the comprehensive characterization of the 12DE region of the *Drosophila melanogaster* genome.

B. Results

The nucleotide sequence of $tRNA_5^{Lys}$ from D. melanogaster was determined by a combination of rapid RNA sequencing techniques (Chapter II). Analysis of nucleotide content (Fig. 32) indicated the presence of pm^1G, pm^2G, pt^6A, $p\psi$, pD, pm^5C, pm^7G, and pm^1A in addition to the four standard nucleotides. In some experiments spots indicating the presence of other modified nucleotides in $tRNA_5^{Lys}$ were seen, but these could not be identified. The sequence of nucleotides 3-69, obtained by the method of Stanley and Vassilenko, is presented in Fig. 33. All the modified nucleotides identified in nucleotide analysis (above) except for pm^1G, pm^2G, and pt^6A were located unambiguously within the $tRNA_5^{Lys}$ sequence, based on chromatographic mobilities of the *pNs in solvents A and B (Fig. 33c) and on location in the tRNA sequence (7,8). Other modified nucleotides found in $tRNA_5^{Lys}$ are discussed below. The nucleosides at positions 1,2, and 70-76 were identified in gel "read-off" experiments (Fig. 35). These experiments also confirmed the identities of nucleosides in many of the positions from 3 to 69. A contaminating $tRNA^{Arg}$ lacking five nucleotides from its 3'-terminus was also analyzed by the gel "read-off" method (Fig. 35a, 39, 40). It is discussed separately below. The sequence from C75 to G65 was determined by "wandering spot" analysis (Fig. 37). A76 was identified by digesting $tRNA_5^{Lys}$, labelled at the 3'end with *pCp with RNase T_2, resulting in transfer of the [^{32}P] phosphoryl group from *pCp to Ap*. The mononucleotide was identified by chromatographic mobilities in solvents A and B relative to nucleoside 3'-phosphate standards, locating the [^{32}P] nucleoside 3'-phosphate by autoradiography.

The nucleotide sequence of $tRNA_5^{Lys}$ is presented in the "cloverleaf" form in Fig. 38.

Figure 32 - Nucleotide analysis of *Drosophila* $tRNA_5^{Lys}$

Nucleotide analysis of $tRNA_5^{Lys}$ was performed as described in Chapter II. Autoradiography of the thin layer plate was for 21 hr. at -20° using Kodak XR-1 "screen" X-ray film.

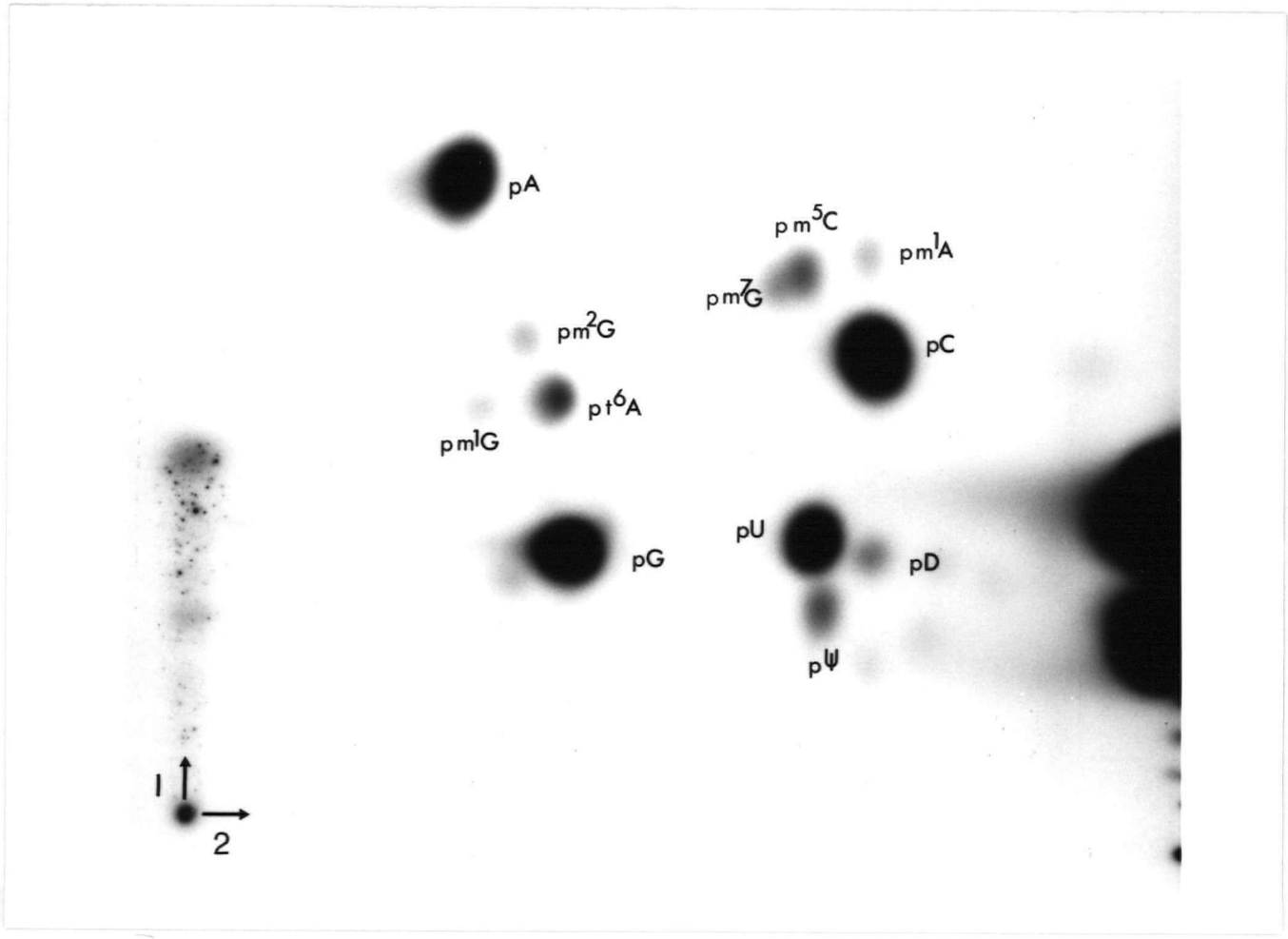


Figure 32

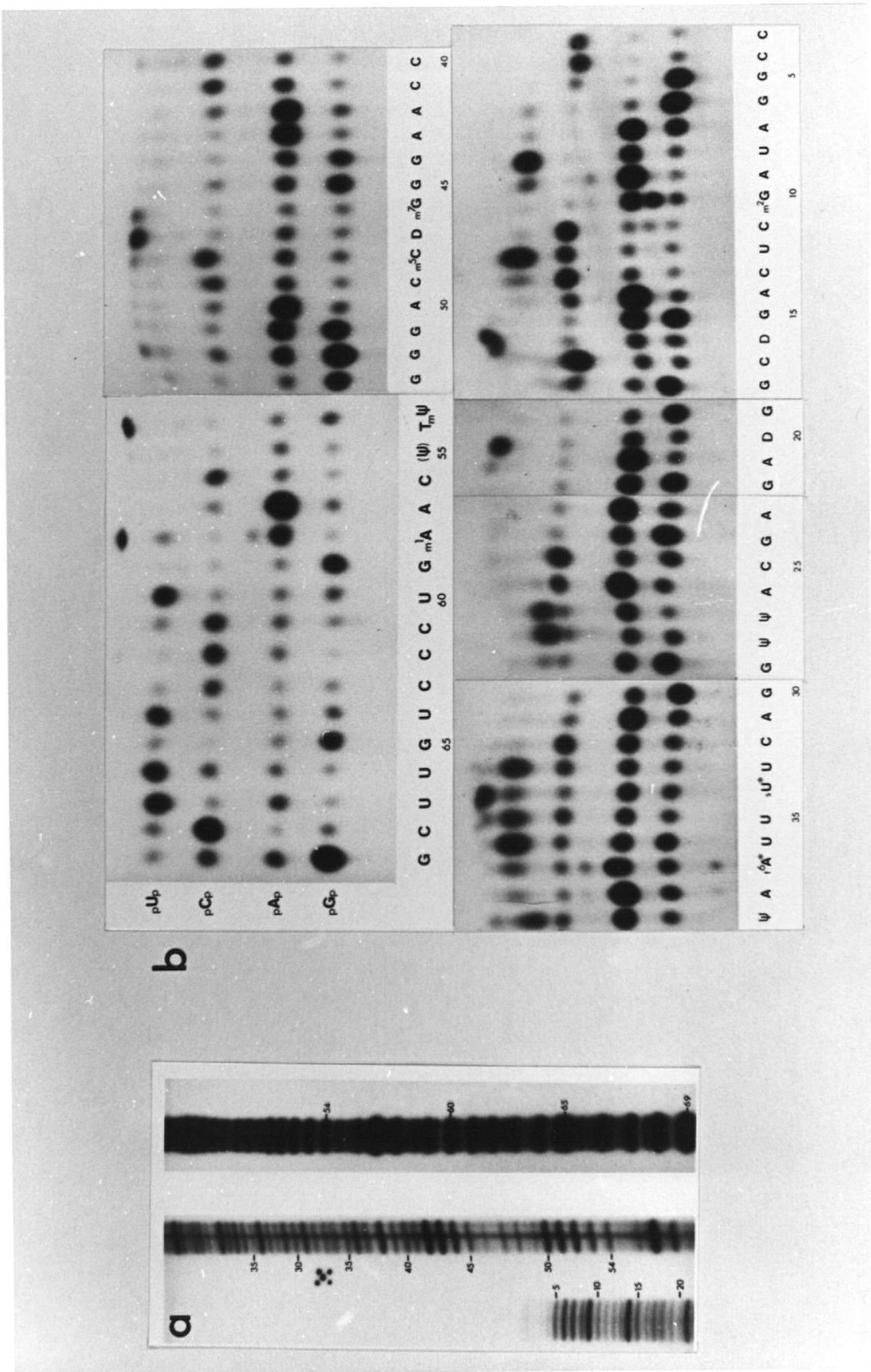
Figure 33 - Stanley/Vassilenko sequence analysis of $tRNA_5^{Lys}$.

(a) Samples of purified $tRNA_5^{Lys}$ (0.8 μg/slot) were partially hydrolyzed in formamide, radiolabelled as described in Chapter II, then separated by polyacrylamide gel electrophoresis as shown. Slots 1 and 2: Electrophoresis was on 12% polyacrylamide gels at a constant power of 30 watts for 3.7 hr. or 1.7 hr., respectively. Slot 3: Electrophoresis was for 2.4 hr. on a 20% polyacrylamide gel at a constant power of 30 watts. Autoradiography was for 20 min. (slots 1,2) or 30 min. (slot 3) using Kodak "no-screen" X-ray film.

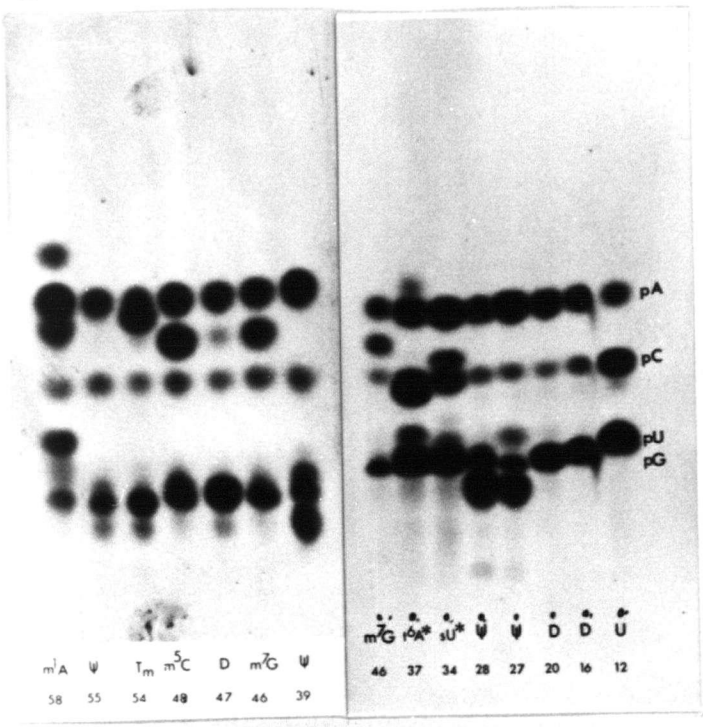
(b) The 5'-terminal nucleotides (*pNp's) of [^{32}P] oligonucleotides separated by polyacrylamide gel electrophoresis as shown in part (a) were analyzed by PEI cellulose chromatography as described in Chapter II. The nucleotides are represented by symbols for the bases, and positions in the tRNA sequence (Fig. 38) indicated below.

(c) The 5'-terminal nucleotides (*pN's) of some [^{32}P] oligonucleotides were analyzed by chromatography on cellulose thin layer plates developing either in solvent A or in solvent B as described in Chapter II. The nucleotides are represented by the symbols for the bases, and positions in the tRNA sequence (Fig. 38) indicated below.

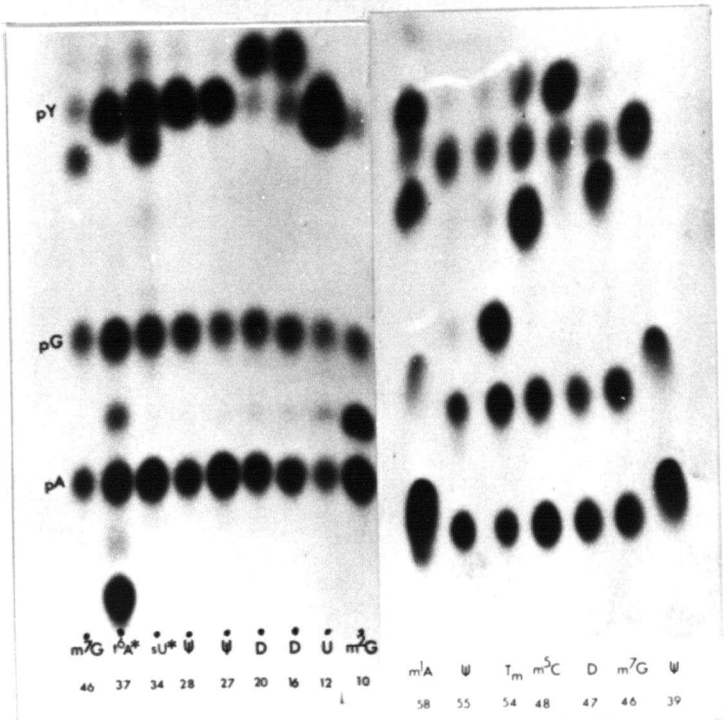
(d) The PEI cellulose plate including positions 54 and 55 of the $tRNA_5^{Lys}$ sequence (part (b) above) was re-exposed to X-ray film for six-fold longer than in part (b). The arrow indicates *pψp.


a
b
pUp
pCp
pAp
pGp
Ψ A m⁶A* U U sU* U C A G G Ψ Ψ A C G A G A D G G C D G A C U C m²G A U A G G C C
G G G A C m⁵C D m²G G G A A C C
G C U U G U C C C U G m¹A A C (Ψ) TmΨ

c


pA
pC
pU
pG
m¹A Ψ Tm m⁵C D m⁷G Ψ
58 55 54 48 47 46 39
m⁷G rA* sU* Ψ Ψ D D U
46 37 34 28 27 20 16 12

Solvent A


pY
pG
pA
m⁷G rA* sU* Ψ Ψ D D U m³G
46 37 34 28 27 20 16 12 10
m¹A Ψ Tm m⁵C D m⁷G Ψ
58 55 54 48 47 46 39

d

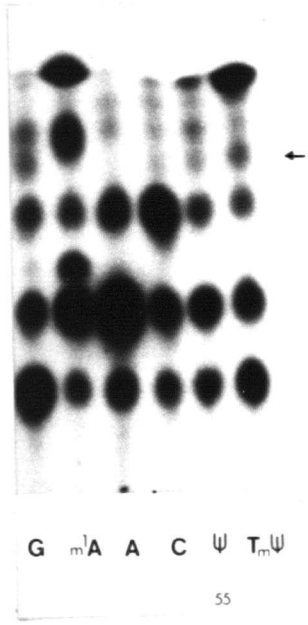

G m¹A A C Ψ TmΨ
55

Figure 34 - Chromatography of a thionucleotide after treatment with CNBr.

The modified nucleotide from position 34 of $tRNA_5^{Lys}$ (-CNBr, arrow) was treated with CNBr, then with HCl for 4 hr. or 22 hr. and chromatographed on cellulose thin layer plates (developing in solvent A), as described in Chapter II. Autoradiography was for 6.5 days at -70° using sensitized Kodak XR-1 "screen" X-ray film (as described in ref. 43).

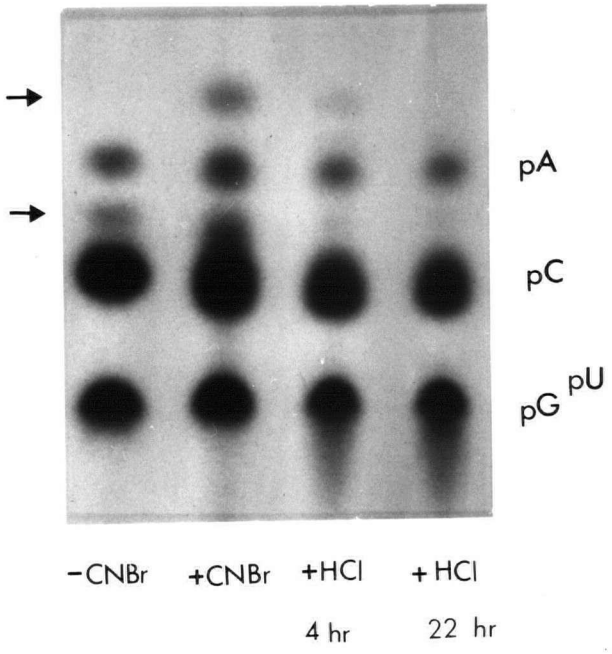


Figure 34.

Figure 35 - Gel "read-off" analysis of [5'-^{32}P]tRNA$_5^{Lys}$ and large oligonucleotides.

(a) Preparation of [5'-^{32}P]RNAs. Purified tRNA$_5^{Lys}$ (2.4 µg) was dephosphorylated as described in Chapter II, then labelled with polynucleotide kinase and [γ-^{32}P]ATP (40 µM, 1200 Ci/mmole). The resulting [^{32}P]RNAs were separated by electrophoresis for 2.7 hr. at 1500V on a denaturing 20% polyacrylamide gel and located by autoradiography for 75 seconds using "no-screen" X-ray film. Band 3 and band 4 include positions 42-76 and 9-41 of the tRNA$_5^{Lys}$ sequence, respectively.

(b) Partial hydrolyses of [5'-^{32}P] tRNA$_5^{Lys}$ with various enzymes were performed in Figure 6(c) above. Electrophoresis was for 2.5 hr. at 1400V on a 20% polyacrylamide gel. RNase T_1 (T), RNase U_2 (U), and control (-E) slots contained 26,000 Cerenkov cpm each, while the RNase Phy I (Phy), RNase A (a), and reference ladder (L) slots contained 52,000 cpm each. Autoradiography was for 3 days at -20° using Kodak "no-screen" X-ray film.

(c) ^{32}P-labelled fragment 4 was digested as above (12000 cpm/slot for T, U, and -E; 24,000 cpm/slot for Phy, A, and L). Electrophoresis was for 2.5 hr. at constant power (34 watts) on a 20% polyacrylamide gel. Autoradiography was for 4 days at -20° using "no-screen" X-ray film.

(d) ^{32}P-labelled fragment 3 was digested, electrophoresed, and exposed to X-ray film as described for fragment 4 in (c) above.

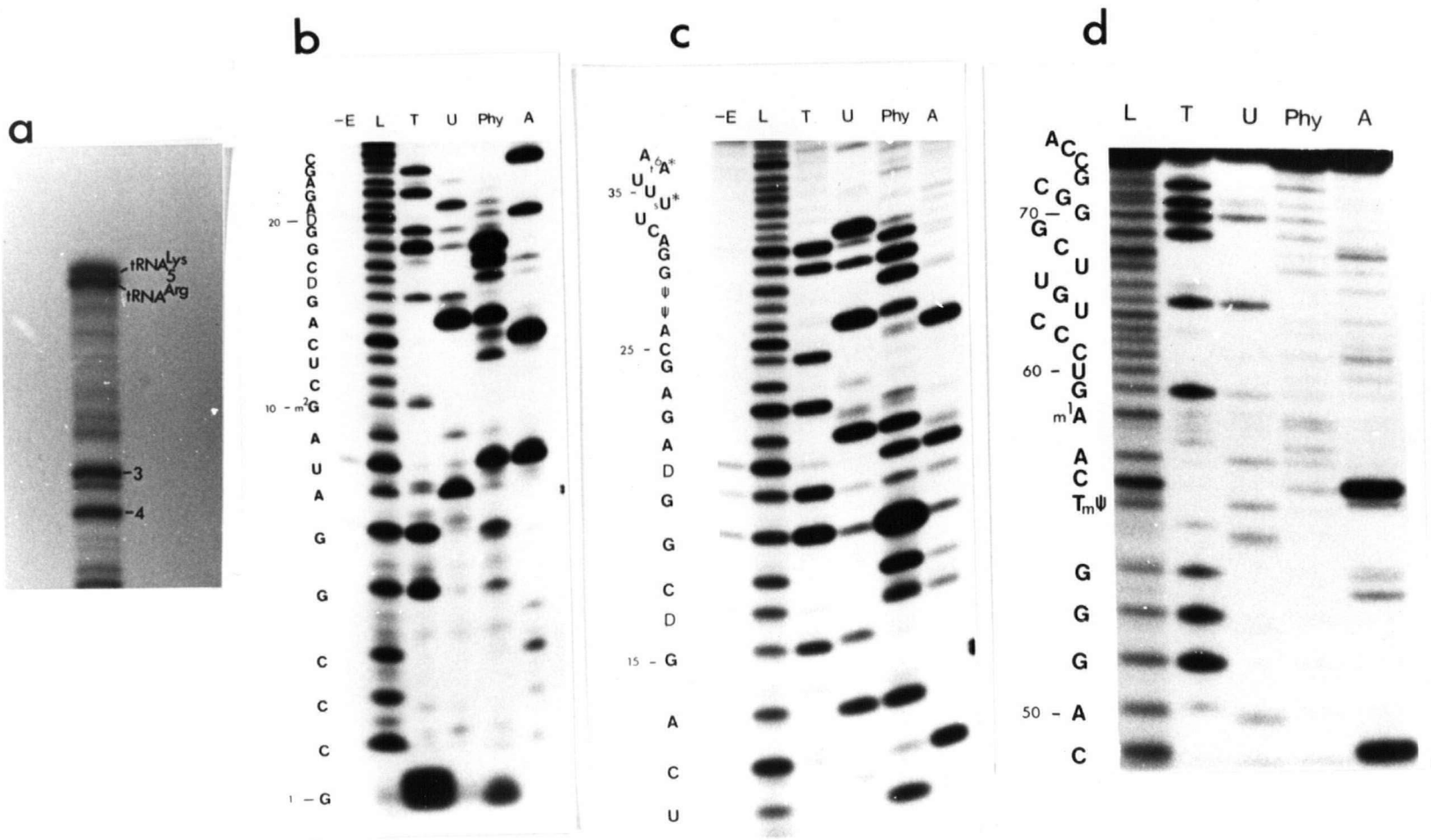

a
tRNA Lys 3
tRNA Arg
-3
-4
b
-E L T U Phy A
c
-E L T U Phy A
d
L T U Phy A

Figure 36 - Gel "read-off" analysis of [3'-^{32}P]tRNA$_5^{Lys}$.

[3'-^{32}P]tRNA$_5^{Lys}$ was prepared as described in Chapter II. Partial hydrolyses with various enzymes were performed as in the legend to Figure 6 (part (c)). Electrophoresis was for 2.5 hr. at 1500V on a 20% polyacrylamide gel. RNase T_1 (T) and RNase U_2 (U) slots contained 5000 Cerenkov cpm while RNase Phy I (Phy), RNase A (A), and formamide ladder (L) slots contained 10,000 cpm. Autoradiography was for 12 days at -20° using Kodak "no-screen" X-ray film.

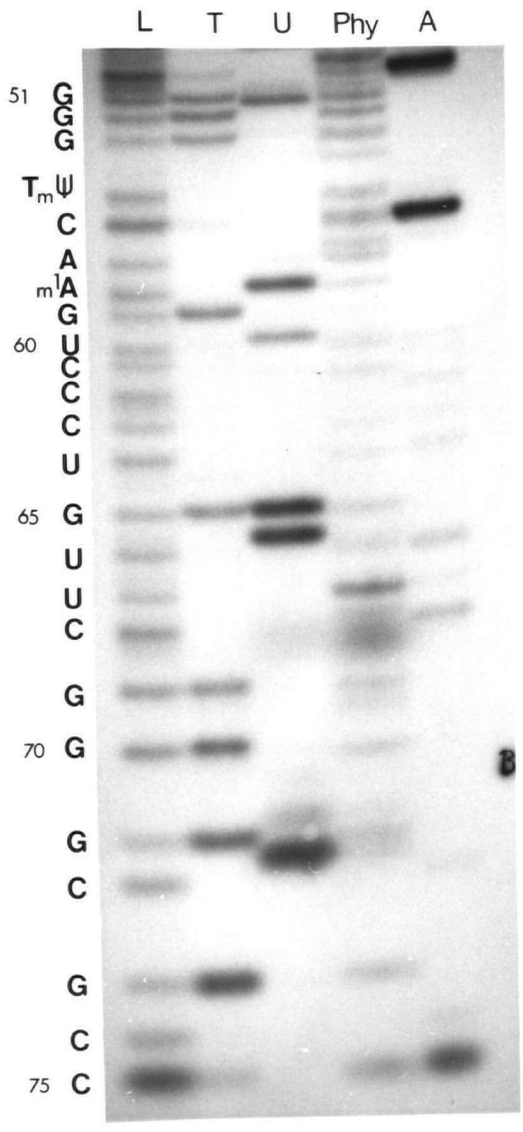


Figure 36.

Figure 37 - "Wandering spot" analysis of [3'-^{32}P]tRNA$_5^{Lys}$

A sample of [3'-^{32}P]tRNA$_5^{Lys}$ (containing 65,000 Cerenkov cpm) was hydrolyzed, electrophoresed on cellulose acetate, transferred to a DEAE-cellulose thin layer plate and developed with homomixture V as described in Chapter II. Autoradiography was for 9 days at -20° using sensitized Kodak XR-1 X-ray film (as described in Methods, ref. 43).

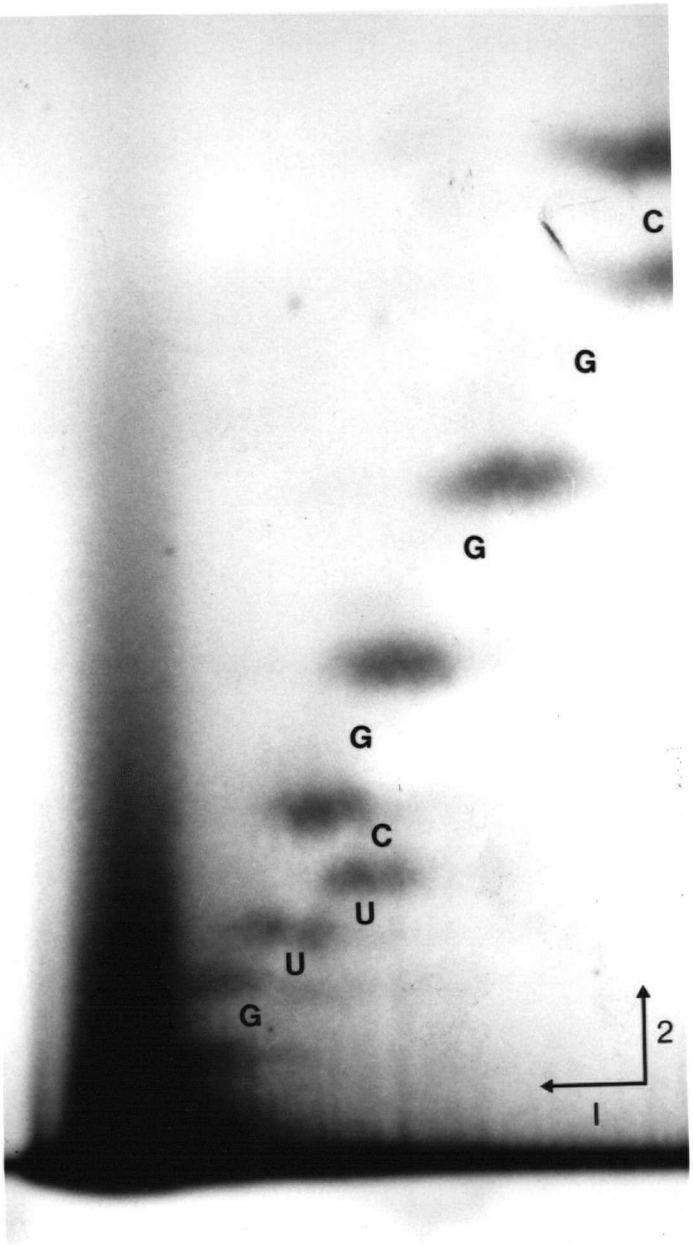
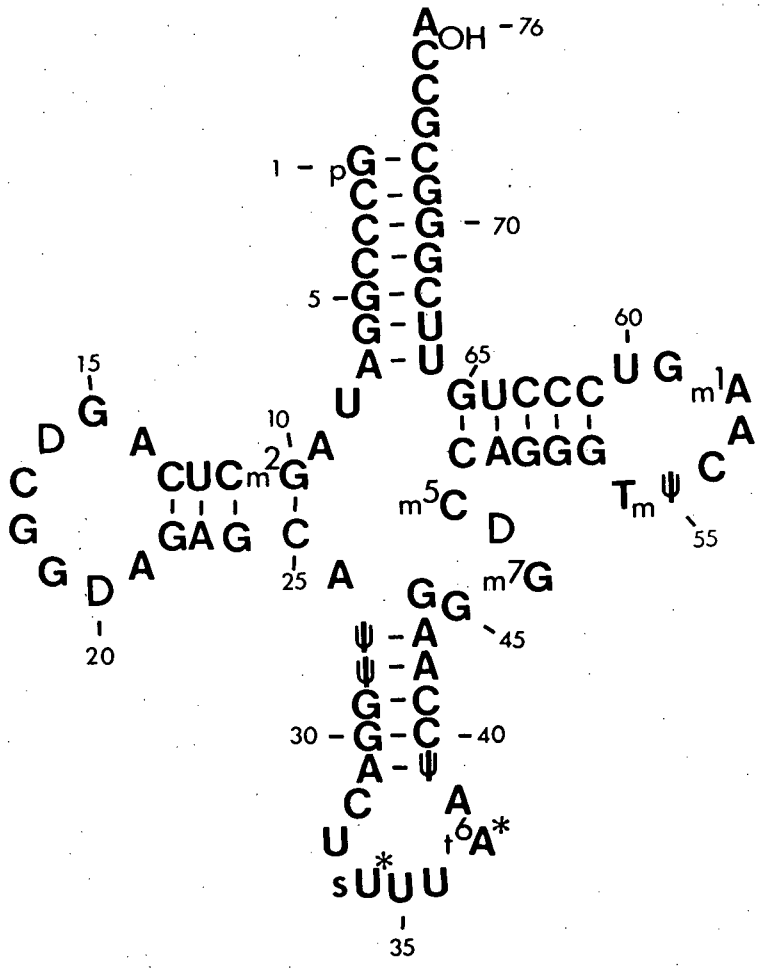


Figure 37.



tRNA$^{Lys}_{5}$

Figure 38 - Cloverleaf structure of tRNA$^{Lys}_{5}$.

Figure 39 - Gel "read-off" analysis of [5'-^{32}P]tRNA$^{Arg}_{CCG}$.

The [5'-^{32}P]tRNA$^{Arg}_{CCG}$ sample was obtained from a tRNA$^{Lys}_{5}$ preparation as shown in Figure 35 (a).

(a) Partial hydrolyses were performed as described in the legend to Figure 6 (part (e)). Control (-E), RNase T_1 (T), and RNase U_2 (U) samples contained 11,000 Cerenkov cpm while RNase Phy I (Phy), RNase A (A), and formamide ladder (L) samples contained twice that amount. Electrophoresis on a 25% polyacrylamide gel was for 3 hr. at constant power (30 watts). Autoradiography was for 6 days at -20° using Kodak "no-screen" X-ray film.

(b) The samples were the same as those used in (a) above. Electrophoresis on a 20% polyacrylamide gel was for 4.25 hr. at a constant power of 34 watts. Autoradiography was for 4 days at -20° using Kodak "no-screen" X-ray film.

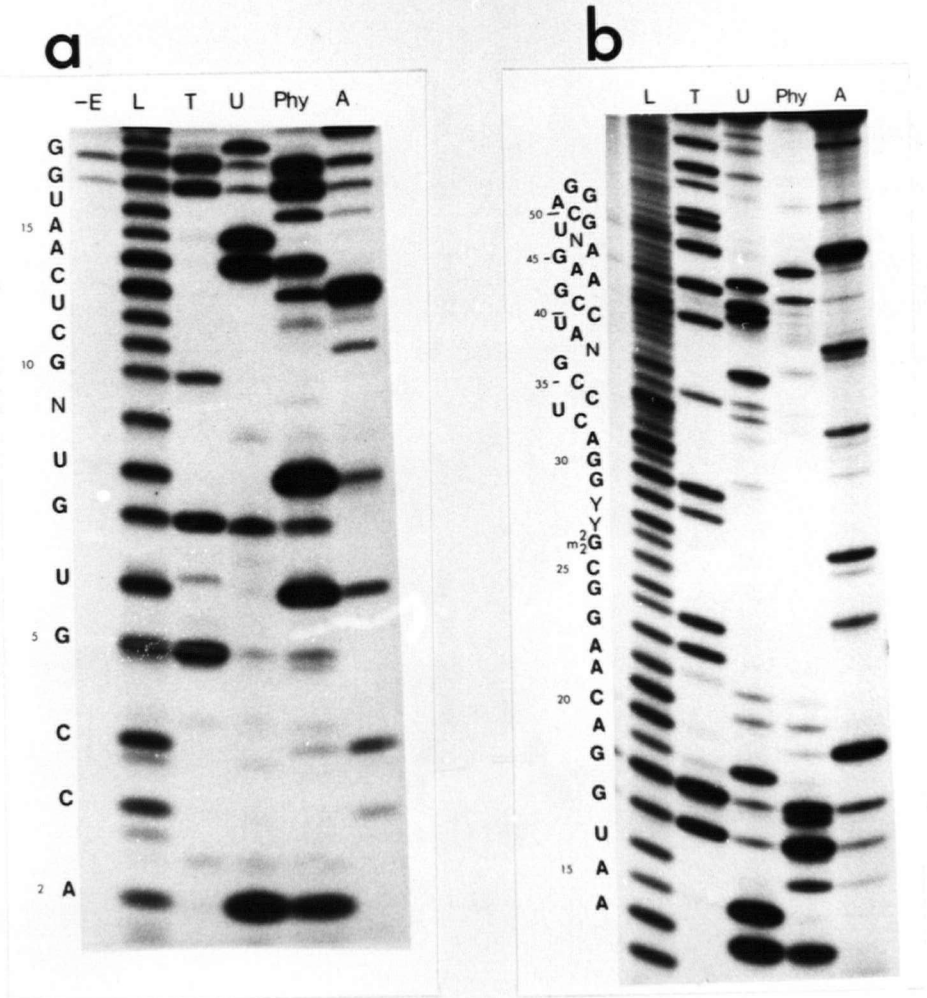

a
-E L T U Phy A
G G U A A C U C G N U G U G C C A
15 10 5 2
b
L T U Phy A
G G C G A A C C N A C C C A G G Y Y G C G G A A C A G G U A A
50 45 40 35 30 25 20 15

Figure 40 - Gel "read-off" analysis of [3'-^{32}P]tRNA$^{Arg}_{CCG}$.

Hydrolyses of [3'-^{32}P]tRNA^{Arg} were performed as described in the legend to Figure 6 (part (c)). Control (-E), RNase T_1 (T), and RNase U_2 (U)) samples contained 9000 Cerenkov cpm/slot while RNase Phy I (Phy), RNase A (A), and formamide ladder (L) samples contained 18,000 cpm/slot. Electrophoresis on a 25% polyacrylamide gel was for 3 hr. at 1500V. Autoradiography was for 44 hr. at -20° using Kodak XR-1 X-ray film.

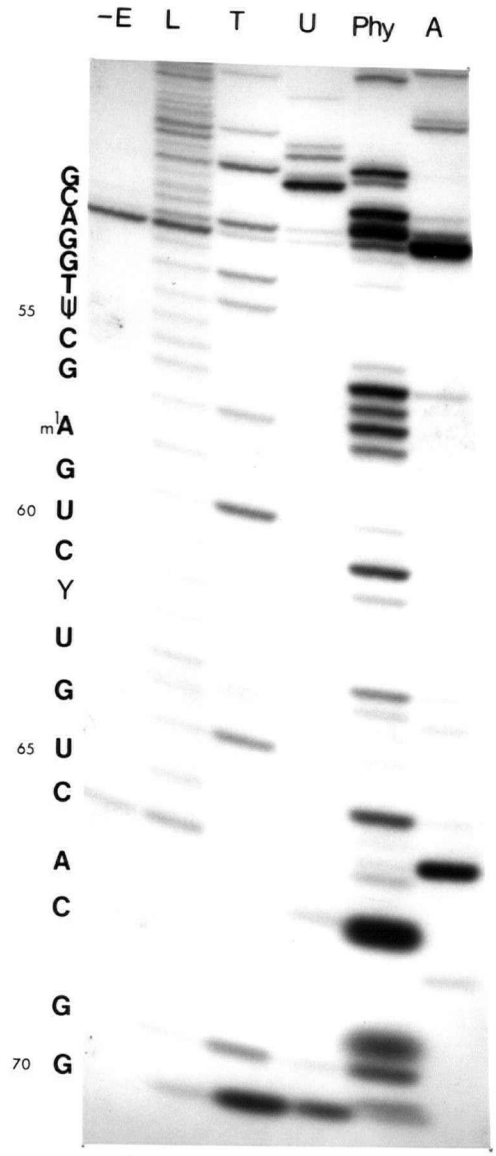


Figure 40.

Modified nucleosides in the $tRNA_5^{Lys}$ sequence other than those identified above include m^2G 10, sU* 34, t^6A* 37, and $T_m\psi$ 54,55 (Fig. 38). These were identified as follows.

m^2G 10: The chromatographic mobilities of the *pN from position 10 were 0.71 in solvent A (relative to pA) and 0.42 in solvent B (relative to pU) compared with published values of ~0.75 and ~0.47 for pm^2G, or ~0.62 and ~0.46 for pm^1G (42). Ribonuclease T_1 (G-specific) cleaves at this position in gel "read-off" experiments (Fig. 35b). Position 10 is in the D stem opposite C25 in the "cloverleaf" structure of Fig. 38. While m^2G is capable of forming Watson-Crick base pairs, m^1G is not. Other eukaryotic lysine tRNAs examined have m^2G in this location (7). The mobility of the *pNp from position 10 (Fig. 33b) corresponds well with that reported for pm^2Gp but not for pm^1Gp in a similar ammonium sulfate solvent (53). On the basis of these observations, m^2G was assigned to position 10.

$T_m\psi$ 54,55: A low background of *pψp was seen in the gap resulting from a 2'-O-methylnucleotide (Fig. 33d), consistent with incomplete ribose-methylation at the sequence $N_m\psi$. The chromatographic mobility of the dinucleoside triphosphate on PEI cellulose in 0.8 M ammonium sulfate was similar to that of pU_mUp from D. melanogaster $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ (Chapter III). The chromatographic mobilities in solvents A and B of the *pN_m derived from the dinucleotide by extensive nuclease P_1 digestion were consistent with those expected for pT_m. The chromatographic behavior of pN's in these solvents is altered in a characteristic way by addition of a 2'-O-methyl group. The mobility of the pN_m is increased by roughly 50% relative to the pN in solvent A, while

in solvent B the pN_m migrates slightly slower than the pN (ref. 42). Relative to the published mobility values for pT (ref. 42), the $*pN_m$ from position 54 behaves as expected for pT_m. A ribonuclease T_2-resistant dinucleotide containing ψ with properties of $T_m\psi p$ has been previously identified in this tRNA (181). Lysine isoacceptor tRNAs from rabbit liver contain $T_m\psi$ (19,180) whereas yeast lysine tRNAs contain $T\psi$ (7). Therefore the sequence at positions 54 and 55 was identified as $T_m\psi$.

sU* 34: The *pNp sample contained at least one modified species, with a mobility greater than pUp (Fig. 33b). One dimensional chromatographs of the pN showed a complex pattern, with at least three species distinct from the four standard nucleotides (Fig. 33c). It has been reported that the amino acid acceptor activity of $tRNA_5^{Lys}$ is abolished by treatment with cyanogen bromide (181), which reacts with sulfur-containing bases (144,145). It was suggested that the CNBr-sensitive species might be a 2-thiouridine derivative in the "wobble" position of the $tRNA_5^{Lys}$ anticodon (181). One modified pN, migrating slightly ahead of pC in solvent A, was modified by CNBr treatment to an acid-sensitive derivative (Fig. 34; ref. 144). In gel "read-off" experiments addition of this nucleotide to a [5'-^{32}P] oligonucleotide results in a large "jump" in the reference ladder, consistent with the presence of a hypermodified nucleotide (Fig. 35c). D. melanogaster contains two major lysine tRNAs ($tRNA_2^{Lys}$ and $tRNA_5^{Lys}$) to read the two lysine codons, AAA and AAG. The sequence of $tRNA_2^{Lys}$ has been reported (104). Its anticodon, CUU, should respond only to AAG. Therefore, $tRNA_5^{Lys}$ must read AAA codons, indicating that N 34 is a derivative of uridine. Based on the chromatographic properties of N 34 as a pNp and as a pN, its properties

in gel "read-off" experiments, and the complexity of the pattern it gives on two-dimensional chromatography (results not shown), its sensitivity to CNBr, and the fact that it should read AAA codons, N 34 is identified as a hypermodified thiouridine derivative. The equivalent lysine tRNAs from yeast and rabbit liver contain mcm^5s^2U in this position. Though N 34 may well be mcm^5s^2U, it could not be positively identified and is presented in Fig. 38 as sU*.

t^6A^* 37: The nucleotide in position 37 is cleaved slowly by RNase Phy I but is resistant to RNases T_1, U_2, or A. The mobility jump observed in the L slot indicates a hypermodified nucleotide (Fig. 35c). This nucleotide gives a complex pattern on one (or two) dimensional thin layer chromatography as the *pN (Fig. 33c). A nucleotide with the chromatographic mobilities of pt^6A was observed in nucleotide analysis (Fig. 32). It was also seen in analysis of *pN's from the gel band corresponding to position 37 (Fig. 33c,d). The predominant form of this nucleotide as a *pNp has a mobility slightly greater than pAp (Fig. 33b), while the *pN has mobilities of 0.67 (relative to pA) in solvent A and 0.1 (relative to pU) in solvent B. The presence of what appears to be pt^6A as a minor component in these experiments suggests that the predominant modified nucleotide is a further modified derivative of pt^6A. Absence of this further modified species in nucleotide analysis (Fig. 32) can be explained by a low affinity of polynucleotide kinase for the nucleoside 3'-phosphate, whereas at the 5'terminus of an oligonucleotide the enzyme labels this nucleotide efficiently. Further experiments confirmed the presence of multiple modified species at position 37, at least one of which is sensitive to CNBr (results not shown). All cytoplasmic eukaryotic tRNAs responding to codons ANN where the identity

of nucleotide 37 is known contain t^6A or a derivative of it (7). In rabbit liver $tRNA_3^{Lys}$, the hypermodified nucleotide 3' to the anticodon has been identified as ms^2t^6A (19,182). Since at least one form of the hypermodified nucleotide in Drosophila $tRNA_5^{Lys}$ is sensitive to CNBr, it is likely that the predominant form of this nucleotide is ms^2t^6A or closely related to it. Due to lack of material, we cannot identify this nucleotide unambiguously, but present it as t^6A*.

m^1G: Though m^1G was identified in nucleotide analysis of $tRNA_5^{Lys}$ here (Fig. 32) and elsewhere (181), it was not located in sequence analysis. A contaminating tRNA fragment representing up to 10% of the RNA was found in our $tRNA_5^{Lys}$ sample (Fig. 35a), despite the multiple steps of purification. The pm^1G seen in nucleotide analysis may come from this contaminant. As m^1G is found in positions 9 and 37 in eukaryotic tRNAs (7), up to 0.2 moles of m^1G per mole of tRNA might be expected, consistent with my analyses.

Sequence analysis of the contaminating tRNA fragment

The sequence of the contaminating tRNA was examined by gel read-off experiments with 5'- or 3'-end-labeled RNA. This contaminant of $tRNA_5^{Lys}$ appears to be $tRNA_{CCG}^{Arg}$ (probably $tRNA_3^{Arg}$, ref. 91) because of its similarity to Drosophila $tRNA_2^{Arg}$. Also, its codon response (CGG) should be complementary to $tRNA_2^{Arg}$, which contains inosine (D. Cribbs, unpublished results) and presumably decodes CGA, CGU, and CGC codons. The partial nucleotide sequence deduced is pGACYGUGUG(m^1G)GCUCAAUGGACAAGGC(m_2^2G)YYGGACU(CC)GNAUCCGAAGANUGC-(A)GGTψCG(m^1A)GU(C)YUGUCACGGU. Despite the fact that the $tRNA_5^{Lys}$ sample was sufficiently pure to use in Stanley and Vassilenko experiments, more ^{32}P was incorporated into the contaminant than $tRNA_5^{Lys}$ by polynucleotide kinase (Fig. 35a). The usual 3'-extended tRNA molecule is not readily phosphorylated by polynucleotide kinase (136), but the contaminant, which lacks five nucleotides from the 3'-end and is therefore 5'-extended, was efficiently labelled. Since the $tRNA^{Arg}$ fragment was also labelled with *pCp by RNA ligase, it contains a 3'-hydroxyl group on the 3'-terminal nucleotide. What enzyme(s) produced this fragment in significant quantity is not clear.

C. Discussion

The nucleotide sequences of $tRNA_5^{Lys}$ and of a $tRNA^{Arg}$ fragment were determined as described above. The $tRNA_5^{Lys}$ sequence shares extensive

homology with other eukaryotic lysine tRNAs. Excluding differences in modification, $tRNA_5^{Lys}$ (sU*UU) is 71% and 95% homologous to the equivalent tRNAs from yeast and rabbit liver, and 82% homologous to $tRNA_2^{Lys}$ (CUU) of D. melanogaster (7). Similarly, $tRNA_2^{Lys}$ is 72% and 100% homologous to the equivalent lysine tRNAs from yeast and rabbit liver (7). The high degree of similarity between equivalent lysine tRNAs from D. melanogaster and rabbit liver (including modifications) indicates extensive conservation of tRNA modifying enzyme activites as well as of tRNA structural genes in multicellular eukaryotes.

D. melanogaster $tRNA_2^{Lys}$ and $tRNA_5^{Lys}$ hybridize in situ to distinct sets of sites on polytene chromosomes (109). None of these sites match the localization at 48F-49A on chromosome 2 originally reported for $tRNA_5^{Lys}$ (108). Whether this site contains genes for $tRNA_{CCG}^{Arg}$ has not been determined. ^{125}I-$tRNA_5^{Lys}$ samples gave grains at the 12DE region of the X chromosome (the major site for $tRNA_4^{Ser}$ and $tRNA_7^{Ser}$ hybridization) when in situ hybridizations were carried at 65° in 2 x SSC, but not at 45° in buffered 70% formamide (conditions where $tRNA_{4,7}^{Ser}$ still hybridize; ref. 109). The grains seen are probably due to cross-hybridization between $tRNA_5^{Lys}$ or $tRNA_{CCG}^{Arg}$ and a similar DNA sequence(s) in the 12DE region. Because the $tRNA_5^{Lys}$ preparation appears to be about 90% pure, it seems more likely that $tRNA_5^{Lys}$ rather than $tRNA_{CCG}^{Arg}$ is the cross-hybridizing species. At any rate, there may well be genes for tRNAs other than $tRNA^{Ser}$ at the 12DE region of the D. melanogaster genome.

A number of recombinant plasmids hybridizing with $tRNA_5^{Lys}$ have been isolated (114). As part of a comprehensive characterisation of lysine

tRNAs and their genes in D. melanogaster, the sequences hybridizing $tRNA_5^{Lys}$ in two plasmids (pDt 39 and pDt 59) were determined. The single gene in pDt 39 has the sequence expected based on the $tRNA_5^{Lys}$ sequence presented here, while the gene in pDt 59 differs at three positions (124). Both genes are transcribed efficiently in vitro in a homologous transcription system (124). Based on the same reasoning as presented in Chapter IV, section B., it seems likely that the $tRNA^{Lys}$ gene differing from $tRNA_5^{Lys}$ is transcribed in vivo as well. However, such a minor species was not identified in the $tRNA_5^{Lys}$ sample analyzed here.

REFERENCES

(1) A. Rich and P.R. Schimmel (1977). Acc. Chem. Res. 10, 385-387.

(2) M.B. Hoagland, P.C. Zamecnik, and M.L. Stephenson (1957). Biochim. Biophys. Acta 24, 215-216.

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

(4) H.G. Zachau, D. Dutting, and H. Feldmann (1966). Hoppe-Seyler's Z. Physiol. Chem. 347, 212-235.

(5) J.T. Madison, G.A. Everett, and H. Kung (1966). Science 153, 531-534.

(6) U.L. RajBhandary, S.H. Chang, A. Stewart, R.D. Faulkner, R.M. Hoskinson, and H.G. Khorana (1967). Proc. Nat. Acad. Sci. USA 57, 751-758.

(7) D.H. Gauss and M. Sprinzl (1981). Nucleic Acids Res. 9 (1), rl-r23.

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

(9) R.P. Singhal and P.A.M. Fallis (1979). Prog. Nucleic Acids Res. Mol. Biol. 23, 227-290.

(10) R. Hall, "The Modified Nucleosides in Nucleic Acids," Columbia University Press, New York and London, 1971.

(11) B. Singer and M. Kroger (1979). Prog. Nucleic Acids Res. Mol. Biol. 23, 151-194.

(12) J.A. McCloskey and S. Nishimura (1977). Acc. Chem. Res. 10, 403-410.

(13) S. Nishimura (1972). Prog. Nucleic Acids Res. Mol. Biol. 12, 49-85.

(14) S. Nishimura, in "Transfer RNA" (ed. S. Altman), M.I.T. Press, Cambridge, Mass., 1978.

(15) D. Soll (1971). Science 173, 293-299.

(16) F.H.C. Crick (1966). J. Mol. Biol. 19, 548-555.

(17) C.T. Caskey, M. Beaudet, and M. Nirenberg (1968). J. Mol. Biol. 37, 99-118.

(18) J.E. Heckman, J. Sarnoff, B. Alzner-DeWeerd, S. Yin, and U.L. RajBhandary (1980). Proc. Nat. Acad. Sci. USA 77, 3159-3163.

(19) M. Raba, K. Limburg, M. Burghagen, J.R. Katze, M. Simsek, J.E. Heckman, U.L. RajBhandary, and H.J. Gross (1979). Eur. J. Biochem. 97, 305-318.

(20) T. Sekiya, K. Takeishi, and T. Ukita (1969). Biochim. Biophys. Acta 182, 411-426.

(21) J.F. Atkins, R.F. Gesteland, B.R. Reid, and C.W. Anderson (1979). Cell 18, 1119-1131.

(22) G.R. Bjork and P.C. Neidhardt (1975). J. Bacteriol. 124, 99-111.

(23) S.H. Kim, G.J. Quigley, F.L. Suddath, A. McPherson, D. Sneden, J.J. Kim, J. Weinzierl, and A. Rich (1973). Science 179, 285-288.

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

(25) J.D. Robertus, J.E. Ladner, J.T. Finch, D. Rhodes, R.S. Brown, B.F.C. Clark, and A. Klug (1974). Nucleic Acids Res. 1, 927-932.

(26) B.F.C. Clark (1977). Prog. Nucleic Acids Res. Mol. Biol. 20, 1-19.

(27) H.T. Wright, P.C. Manor, K. Beurling, R.L. Karpel, and J.R. Fresco. In "Transfer RNA: Structure, Properties, and Recognition" (ed. P.R. Schimmel, D. Soll, and J.N. Abelson), Cold Spring Harbor Laboratory, 1979.

(28) R.W. Schevitz, A.D. Podjarny, N. Krishnamachari, J.J. Hughes, P.B. Sigler, and J.L. Sussman (1979). Nature 278, 188-190.

(29) "Procedures in Nucleic Acid Research," Vol. 2 (ed. G.L. Cantoni and D.R. Davies), Harper and Row, 1971.

(30) I. Gillam, S. Millward, D. Blew, M. von Tigerstrom, E. Wimmer, and G.M. Tener (1967). Biochemistry 6, 3043-3056.

(31) I. Gillam, D. Blew, R.C. Warrington, M. von Tigerstrom, and G.M. Tener (1968). Biochemistry 7, 3459-3468.

(32) S. Nishimura, F. Harada, U. Narushima, and T. Seno (1967). Biochim. Biophys. Acta 142, 133-138.

(33) W.M. Holmes, R.E. Hurd, B.R. Reid, R.A. Rimerman, and G.W. Hatfield (1975). Proc. Nat. Acad. Sci. USA 72, 1068-1071.

(34) R.L. Pearson, J.F. Weiss, and A.D. Kelmers (1971). Biochim. Biophys. Acta 228, 770-774.

(35) R.P. Singhal (1974). Eur. J. Biochem. 43, 245-252.

(36) H. Grosjean, C. Takada, and J. Petre (1973). Biochem. Biophys. Res. Comm. 53, 882-893.

(37) R. Salomon, S. Fuchs, A. Aharonov, D. Giveon, and U.Z. Littauer (1975). Biochemistry 14, 4046-4050.

(38) U.L. RajBhandary, R.D. Faulkner, and A. Stuart (1968). J. Biol. Chem. 243, 575-583.

(39) E. Kubli, T. Schmidt, and A.H. Egg, in "Transfer RNA: Biological Aspects" (ed. D. Soll, J.N. Abelson, and P.R. Schimmel), Cold Spring Harbor Laboratory, 1980.

(40) A. Fradin, H. Gruhl, and H. Feldmann (1975). FEBS Lett. 50, 185-189.

(41) E. Randerath, R.C. Gupta, H.P. Morris, and K. Randerath (1980). Biochemistry 19, 3476-3483.

(42) M. Silberklang, A.M. Gillum, and U.L. RajBhandary (1979). Meth. in Enzymol. 54, Part G, 58-109.

(43) D.L. Cribbs (1979). M.Sc. thesis, University of British Columbia.

(44) R.W. Holley (1968). Prog. Nucleic Acid Res. Mol. Biol. 8, 37-47.

(45) F. Sanger, G.G. Brownlee, and B.G. Barrell (1965). J. Mol. Biol. 13, 373-398.

(46) K. Randerath, R.C. Gupta, and E. Randerath, in "Transfer RNA: Structure, Properties, and Recognition" (ed. P.R. Schimmel, D. Soll, and J.N. Abelson), Cold Spring Harbor Laboratory, 1979.

(47) E. Jay, R. Bambara, R. Padmanabhan, and R. Wu (1974). Nucleic Acids Res. 1, 331-353.

(48) R.E. Lockard, B. Alzner-Deweerd, J.E. Heckman, J. MacGee, M.W. Tabor, and U.L. RajBhandary (1978). Nucleic Acids Res. 5, 37-56.

(49) H. Donis-Keller, A.M. Maxam, and W. Gilbert (1977). Nucleic Acids Res. 4, 2527-2538.

(50) A. Simoncsits, G.G. Brownlee, R.S. Brown, J.R. Rubin, and H. Guilley (1977). Nature 269, 833-836.

(51) D.A. Peattie (1979). Proc. Nat. Acad. Sci. USA 76, 1760-1764.

(52) J. Stanley and S. Vassilenko (1978). Nature 274, 87-89.

(53) R.C. Gupta and K. Randerath (1979). Nucleic Acids Res. 6, 3443-3458.

(54) Y. Tanaka, T.A. Dyer, and G.G. Brownlee (1980). Nucleic Acids Res. 8, 1259-1272.

(55) H.P. Agrawal, K. Randerath, and E. Randerath (1981). Nucleic Acids Res. 9, 2535-2541.

(56) H.M. Sprouse, M. Kashdan, L. Otis, and B. Dudock (1981). Nucleic Acids Res. 9, 2543-2547.

(57) A. Rafalski, J. Kohli, P. Agris, and D. Soll (1979). Nucleic Acids Res. 6, 2683-2695.

(58) J. Abelson In "Transfer RNA: Biological Aspects" (ed. D. Soll, J.N. Abelson, and P.R. Schimmel), Cold Spring Harbor Laboratory, 1980.

(59) G.P. Mazzara and W.H. McClain. In "Transfer RNA: Biological Aspects" (ed. D. Soll, J.N. Abelson, and P.R. Schimmel), Cold Spring Harbor Laboratory, 1980.

(60) E. Lund and J.E. Dahlberg (1977). Cell 11, 247-262.

(61) E. Lund, J.E. Dahlberg, L. Lindahl, S.R. Jaskunis, P.P. Dennis, and M. Nomura (1976). Cell 7, 165-177.

(62) J. Rossi, J. Egan, M.L. Berman, and A. Landy. In "Transfer RNA: Biological Aspects" (ed. D. Soll, J.N. Abelson, and P.R. Schimmel), Cold Spring Harbor Laboratory, 1980.

(63) S. Altman and J.D. Smith (1971). Nature New Biol. 233, 35-39.

(64) H.G. Khorana, K.L. Agarwal, P. Besmer, H. Buchi, M.H. Caruthers, P.J. Cashion, M. Fridkin, E. Jay, K. Kleppe, R. Kleppe, A. Kumar, P.C. Loewen, R.C. Miller, K. Minamoto, A. Panet, U.L. RajBhandary, B. Ramamoorthy, T. Sekiya, T. Takeya, and J.H. van de Sande (1976). J. Biol. Chem. 251, 565-570.

(65) M.V. Olson, G.S. Page, A. Sentenac, K. Loughney, J. Kurjan, J. Benditt, and B.D. Hall. In "Transfer RNA: Biological Aspects" (ed. D. Soll, J.N. Abelson, and P.R. Schimmel), Cold Spring Harbor Laboratory, 1980.

(66) M.V. Olson, B.D. Hall, J.R. Cameron, and R.W. Davis (1979). J. Mol. Biol. 127, 285-295.

(67) M.V. Olson, G.S. Page, A. Sentenac, P.W. Piper, M. Worthington, R.B. Weiss, and B.D. Hall (1981). Nature 291, 464-469.

(68) S.G. Clarkson, M.L. Birnstiel, and I.F. Purdom (1973). J. Mol. Biol. 79, 411-429.

(69) S.G. Clarkson, M.L. Birnstiel, and V. Serra (1973). J. Mol. Biol. 79, 391-410.

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

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

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

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

(74) P. Arcari and G.G. Brownlee (1980). Nucleic Acids Res. 8, 5207-5212.

(75) M.H.L. de Bruijn, P.H. Schreier, I.C. Eperon, B.G. Barrell, E.Y. Chen, P.W. Armstrong, J.F.H. Wong, and B.A. Roe (1980). Nucleic Acids Res. 8, 5213-5222.

(76) B.G. Barrell, A.T. Bankier, and J. Drouin (1979). Nature 282, 189-194.

(77) G. Macino, G. Coruzzi, F.G. Nobrega, M. Li, and A. Tzagoloff (1979). Proc. Nat. Acad. Sci. USA 76, 3784-3785.

(78) S. Anderson, A.T. Bankier, B.G. Barrell, M.H.L. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J.H. Smith, R. Staden, and I.G. Young (1981). Nature 290, 457-465.

(79) D. Ojala, J. Montoya, and G. Attardi (1981). Nature 290, 470-474.

(80) S.M. Tilghman, D.C. Tiemeier, J.G. Seidman, B.M. Peterlin, M. Sullivan, J.V. Maizel, and P. Leder (1978). Proc. Nat. Acad. Sci. USA 75, 725-727.

(81) R. Breathnach, J.L. Mandel, and P. Chambon (1977). Nature 270, 314-319.

(82) P.K. Wellauer and I.B. Dawid (1977). Cell 10, 193-212.

(83) R.L. White and D.S. Hogness (1977). Cell 10, 177-192.

(84) H.M. Goodman, M.V. Olson, and B.D. Hall (1977). Proc. Nat. Acad. Sci. USA 74, 5453-5457.

(85) P. Valenzuela, A. Venegas, F. Weinberg, R. Bishop, and W.J. Rutter (1978). Proc. Nat. Acad. Sci. USA 75, 190-194.

(86) A. Venegas, M. Quiroga, J. Zaldivar, W.J. Rutter, and P. Valenzuela (1979). J. Biol. Bhem. 254, 12,306-12,309.

(87) R.R. Robinson and N. Davidson (1980). Cell 23, 251-259.

(88) B. Wittig and S. Wittig (1979). Cell 18, 1173-1183.

(89) W. Koch, K. Edwards, and H. Kossel (1981). Cell 25, 203-213.

(90) R.A. Young, R. Macklis, and J.A. Steitz (1979). J. Biol. Chem. 254, 3264-3271.

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

(92) B.N. White, G.M. Tener, J. Holden, and D.T. Suzuki (1973). J. Mol. Biol. 74, 635-651.

(93) M. Altwegg and E. Kubli (1980). Nucleic Acids Res. 8, 3259-3262.

(94) M.A. Wosnick and B.N. White (1977). Nucleic Acids Res. 4, 3919-3930.

(95) K.B. Jacobsen (1971). Nature New Biol. 231, 17-19.

(96) D.R. Twardzic, E.H. Grell, and K.B. Jacobsen (1971). J. Mol. Biol. 57, 231-245.

(97) D. Mischke, P. Kloetzel, and M. Schwochau (1975). Nature 255, 79-80.

(98) R. Dunn, W.R. Addison, I.C. Gillam, and G.M. Tener (1978). Can. J. Biochem. 56, 618-623.

(99) Robert J. Dunn (1976). Ph.D. thesis, University of British Columbia.

(100 W.R. Addison, unpublished results.

(101) W.R. Addison, I.C. Gillam, and G.M. Tener (1982). J. Biol. Chem., in press.

(102) B.N. White, R. Dunn, I. Gillam, G.M. Tener, D.J. Armstrong, F. Skoog, C.R. Frihart, and N.J. Leonard (1975). J. Biol. Chem. 250, 515-521.

(103) M. Altwegg and E. Kubli (1980). Nucleic Acids Res. 8, 215-223.

(104) S. Silverman, I.C. Gillam, G.M. Tener, and D. Soll (1979). Nucleic Acids Res. 6, 435-442.

(105) S. Silverman, J. Heckman, G.J. Cowling, A.D. Delaney, R.J. Dunn, I.C. Gillam, G.M. Tener, D. Soll, and U.L. RajBhandary (1979). Nucleic Acids Res. 6, 421-433.

(106) M. Altwegg and E. Kubli (1979). Nucleic Acids Res. 7, 93-105.

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

(108) T.A. Grigliatti, B.N. White, G.M. Tener, T.C. Kaufman, and D.T. Suzuki (1974). Proc. Nat. Acad. Sci. USA 71, 3527-3531.

(109) S. Hayashi, I.C. Gillam, A.D. Delaney, R. Dunn, G.M. Tener, T.A. Grigliatti, and D.T. Suzuki (1980). Chromosoma 76, 65-84.

(110) S. Hayashi, W.R. Addison, I.C. Gillam, T.A. Grigliatti, and G.M. Tener (1981). Chromosoma 82, 385-397.

(111) E. Kubli, T. Schmidt, and A.H. Egg. In "Transfer RNA: Biological Aspects" (ed. D. Soll, J.N. Abelson, and P.R. Schimmel), Cold Spring Harbor Laboratory, 1980.

(112) R.T. Elder, O.C. Uhlenbeck, and P. Szabo. In "Transfer RNA: Biological Aspects" (ed. D. Soll, J.N. Abelson, and P.R. Schimmel), Cold Spring Harbor Laboratory, 1980.

(113) R. Dunn, S. Hayashi, I.C. Gillam, A.D. Delaney, G.M. Tener, T.A. Grigliatti, T.C. Kaufman, and D.T. Suzuki (1979). J. Mol. Biol. 128, 277-287.

(114) R. Dunn, A.D. Delaney, I.C. Gillam, S. Hayashi, G.M. Tener, T. Grigliatti, V. Misra, M.G. Spurr, D.M. Taylor, and R.C. Miller, Jr. (1979). Gene 7, 197-215.

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

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

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

(118) J.P. Gergen, J.Y. Loewenberg, and P.C. Wensink (1981). J. Mol.. Biol. 147, 475-499.

(119) H.A. Hosbach, M. Silberklang, and B.J. McCarthy (1980). Cell 21, 169-178.

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

(121) S. Sharp, D. De Franco, M. Silberklang, H. Hosbach, J.P. Gergen, P.C. Wensink, R. Dudler, E. Kubli, and D. Soll (1981). Manuscript submitted.

(122) W.R. Addison, C.R. Astell, A.D. Delaney, I.C. Gillam, S. Hayashi, R.C. Miller, B. Rajput, M. Smith, D.M. Taylor, and G.M. Tener (1982). J. Biol. Chem., in press.

(123) D. De Franco, O. Schmidt, and D. Soll (1980). Proc. Nat. Acad. Sci. USA 77, 3365-3368.

(124) D. Soll, personal communication.

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

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

(127) S. Silverman, O. Schmidt. D. Soll, and B. Hovemann (1979). J. Biol. Chem. 254, 10290-10294.

(128) E.M. DeRobertis and M.V. Olson (1979). Nature 278, 137-143.

(129) T. Dingermann, S. Sharp, B. Appel, D. DeFranco, S. Mount, R. Heiermann, O. Pongs, and D. Soll (1981). Nucleic Acids Res. 9, 3907-3918.

(130) K.U. Sprague, D. Larsen, D. Morton (1980). Cell 22, 171-178.

(131) I.M. Glynn and J.B. Chappell (1964). Biochem. J. 90, 147-149.

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

(133) M. Fujimoto, A. Kuninaka, and H. Yoshino (1974). Agr. Biol. Chem. 38, 1555-1561.

(134) A.G. Bruce and O.C. Uhlenbeck (1978). Nucleic Acids Res. 5, 3665-3677.

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

(136) J.R. Lillehaug and K. Kleppe (1977). Nucleic Acids Res. 4, 373-379.

(137) J.-P. Garel, R.L. Garber, and M.A.Q. Siddiqui (1977). Biochemistry 16, 3618-3624.

(138) M.I. Simon and H. Van Vunakis (1964). Arch. Biochem. Biophys. 105, 197-206.

(139) L.H. Schulman (1972). Proc. Nat. Acad. Sci. USA 69, 3594-3597.

(140) J.A. Haines, C.B. Reese, and A.R. Todd (1962). J. Chem. Soc. 5281-5288.

(141) E.M. Southern and A.R. Mitchell (1971). Biochem. J. 123, 613-617.

(142) Y. Yamada and H. Ishikura (1975). Biochim. Biophys. Acta 402, 285-287.

(143) K. Randerath, R.C. Gupta, and E. Randerath (1980). Meth. in. Enzymol., vol. 65, 638-680.

(144) M. Saneyoshi and S. Nishimura (1970). Biochim. Biophys. Acta 204, 389-399.

(145) M. Saneyoshi and S. Nishimura (1971). Biochim. Biophys. Acta 246, 123-131.

(146) B.G. Barrell, in "Procedures in Nucleic Acid Research," Vol. 2 (ed. G.L. Cantoni and D.R. Davies), pp. 751-795. Harper and Row, New York, 1971.

(147) E.M. Southern (1979). Anal. Biochem. 100, 319-323.

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

(149) K. Murray and N.E. Murray (1975). J. Mol. Biol. 98, 551-564.

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

(151) H.O. Smith and M.L. Birnstiel (1976). Nucleic Acids Res. 3, 2387-2398.

(152) A.M. Maxam and W. Gilbert (1980). Methods in Enzymol., vol. 65, pp. 499-559.

(153) Deleted in proof.

(154) D.L. Lindsley and E.H. Grell, "Genetic Variations of Drosophila melanogaster," Carnegie Institution, Washington, D.C., 1972.

(155) S. Hayashi, unpublished results.

(156) B. Rajput, unpublished results.

(157) D.M. Brown, A. Todd, and S. Varadarajan (1956). J. Chem. Soc. 2384.

(158) G.W. Rushizky and H.A. Sober (1968). Prog. Nucleic Acid Res. Mol. Biol. 8, 171-207.

(159) E. Randerath, C.-T. Yu, and K. Randerath (1972). Anal. Biochem. 48, 172-198.

(160) F.M. Richards and H.W. Wyckoff, in "The Enzymes" (ed. P.D. Boyer), Academic Press, New York and London, 1971.

(161) P.R. Schimmel and D. Soll (1979). Ann. Rev. Biochem. 48, 601-648.

(162) A. Rich and P.R. Schimmel (1977). Nucleic Acids Res. 4, 1649-1665.

(163) S.A. Endow and R.J. Roberts (1977). J. Mol. Biol. 112, 521-529.

(164) B. Rajput, D. DeMille, R.C. Miller, and G.B. Spiegelman, personal communication.

(165) B.-I. Ono, N. Wills, J.W. Stewart, R.F. Gesteland, and F. Sherman (1981). J. Mol. Biol. 150, 361-373.

(166) M.C. Brandriss, J.W. Stewart, F. Sherman, and D. Botstein (1976). J. Mol. Biol. 102, 467-476.

(167) A. Hinnen, J.B. Hicks, and G.R. Fink (1978). Proc. Nat. Acad. Sci. USA 75, 1929-1933.

(168) S. Scherer and R.W. Davis (1980). Science 209, 1380-1384.

(169) H.L. Klein and T.D. Petes (1981). Nature 289, 144-148.

(170) J.A. Jackson and G.R. Fink (1981). Nature 292, 306-311.

(171) J.F. Ernst, J.W. Stewart, and F. Sherman (1981). Proc. Nat. Acad. Sci. USA 78, 6334-6338.

(172) H. Hottinger and U. Leupold (1981). Current Genetics 3, 133-143.

(173) J.L. Slightom, A.E. Blechl, and O. Smithies (1980). Cell 21, 627-638.

(174) S.-h. Shen, J.L. Slightom, and O. Smithies (1981). Cell 26, 191-203.

(175) D. Baltimore (1981). Cell 24, 592-594.

(176) H. Rogg, P. Muller, and M. Staehelin (1975). Eur. J. Biochem. 53, 115-127.

(177) R.L. Sinsheimer (1977). Ann. Rev. Biochem. 46, 415-438.

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

(179) E.H. Birkenmeier, D.D. Brown, and E. Jordan (1978). Cell 15, 1077-1086.

(180) H.J. Gross, M. Simsek, M. Raba, K. Limburg, J. Heckman, and U.L. RajBhandary (1974). Nucleic Acids Res. 1, 35-44.

(181) B.N. White (1975). Biochim. Biophys. Acta 395, 322-328.

(182) A. Yamaizumi, S. Nishimura, K. Limburg, M. Raba, H.J. Gross, P.F. Crain, and J.A. McCloskey (1979). J. Am. Chem. Soc. 101, 2224-2225.