1979

The Primary Structure of Euglena Gracilis Cytoplasmic Phenylalanine Transfer Rna.

Charles Kasriel Brum
Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_disstheses/3324

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.
BRUM, CHARLES KABRIEL
THE PRIMARY STRUCTURE OF EUGLENA GRACILIS CYTOPLASMIC PHENYLALANINE TRANSFER RNA.

THE LOUISIANA STATE UNIVERSITY AND AGRICULTURAL AND MECHANICAL COL., PH.D., 1979
THE PRIMARY STRUCTURE OF EUGLENA GRACILIS

CYTOPLASMIC PHENYLALANINE TRANSFER RNA

A DISSERTATION

Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in

The Department of Biochemistry

by

Charles Kasriel Brum
B.S., University of New Orleans, New Orleans, 1974
May, 1979
ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr. Simon H. Chang for his constant guidance and collaboration throughout the course of this project. I would also like to thank Dr. K. Randerath of Baylor Medical School and Dr. U. L. RajBhandary of the Massachusetts Institute of Technology in whose laboratories I learned many of the techniques utilized in our research. I would like to express my appreciation to Dr. W. E. Barnett and Dr. L. Hecker for providing purified \textit{Euglena gracilis} chloroplastic and cytoplasmic tRNA$^{\text{Phe}}$. I am indebted to Amanda Gillum for guidance while I was at MIT. I would especially like to thank my friend, James Schnabel, who sequenced the pancreatic RNase generated oligonucleotides of \textit{Euglena gracilis} cytoplasmic tRNA, for his support and collaboration in this project.

I would like to thank the members of my dissertation committee for the time they took to evaluate my dissertation. I would like to thank the many friends I have made during my stay at LSU whose friendship I will always remember. I would like to express by appreciation to Ella Kelley for her excellent typing.

I am indebted to my brother, Morris, and my mother, Mrs. H. Brum for their moral support. I will always cherish the memory of my late father, Henry Brum, who, while himself denied the opportunity to obtain an education, always encouraged me to get one.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xi</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>xiii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>I. Biological Functions of tRNA</td>
<td>1</td>
</tr>
<tr>
<td>II. General Aspects of RNA Sequence Determination</td>
<td>4</td>
</tr>
<tr>
<td>III. The General Structure of tRNA</td>
<td>7</td>
</tr>
<tr>
<td>IV. Structure-Function Relationships in tRNA</td>
<td>14</td>
</tr>
<tr>
<td>V. The Evolution of tRNA</td>
<td>17</td>
</tr>
<tr>
<td>VI. Rationale for this Study</td>
<td>20</td>
</tr>
<tr>
<td>MATERIALS AND METHODS</td>
<td>24</td>
</tr>
<tr>
<td>I. Materials</td>
<td>24</td>
</tr>
<tr>
<td>A. Specialized Materials for Autoradiography, Chromatography, High Voltage Electrophoresis, and Radiation Safety</td>
<td>24</td>
</tr>
<tr>
<td>B. Enzymes</td>
<td>25</td>
</tr>
<tr>
<td>C. Radioactive Materials and Fluors</td>
<td>26</td>
</tr>
<tr>
<td>D. Specialized Reagents and tRNA Samples</td>
<td>27</td>
</tr>
<tr>
<td>E. Instrumentation and Equipment</td>
<td>28</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>II. Methods</td>
<td>29</td>
</tr>
<tr>
<td>A. General Methods</td>
<td>29</td>
</tr>
<tr>
<td>B. Methods for Base Composition and Sequence Analysis of RNA by Chemical Tritium Labelling</td>
<td>32</td>
</tr>
<tr>
<td>C. Methods for Sequencing Using 5'-End Group Labelling with [32p]</td>
<td>37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RESULTS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Base Composition Analysis of <em>Euglena gracilis</em> Cytoplasmic tRNAPhe by Chemical Tritium Labelling</td>
<td>58</td>
</tr>
<tr>
<td>II. Base Composition Analysis of T₁ RNase Generated Oligonucleotides by Chemical Tritium Labelling</td>
<td>65</td>
</tr>
<tr>
<td>III. Sequencing of a T₁ RNase Fragment Utilizing Chemical Tritium Labelling</td>
<td>79</td>
</tr>
<tr>
<td>IV. Sequencing of Oligonucleotides Utilizing 5'-End Group [32p]-Labelling</td>
<td>85</td>
</tr>
<tr>
<td>A. Base Specific Digests, 5'-End Group [32p]-Labelling with Polynucleotide Kinase, and Fingerprinting</td>
<td>85</td>
</tr>
<tr>
<td>B. 5'-End Group Determination</td>
<td>90</td>
</tr>
<tr>
<td>C. Sequencing by Analysis of 5'-[32p]-Labelled Oligonucleotides</td>
<td>99</td>
</tr>
<tr>
<td>D. Sequencing of Oligonucleotides Containing &quot;Problem&quot; Modified Bases</td>
<td>129</td>
</tr>
<tr>
<td>E. Creation of Overlap Sequences</td>
<td>157</td>
</tr>
<tr>
<td>F. The Nucleotide Sequence of <em>Euglena gracilis</em> Cytoplasmic tRNAPhe</td>
<td>165</td>
</tr>
</tbody>
</table>

DISCUSSION                                                                 | 171  |
REFERENCES .................................................. 184
VITA ............................................................. 192
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Comparison of the Nucleoside Composition of Euglena tRNAs\textsuperscript{phe} by Uziel's and Randerath's Methods</td>
<td>59</td>
</tr>
<tr>
<td>2.</td>
<td>Nucleoside Composition of tRNAs\textsuperscript{phe} of Euglena Cytoplasm, Beef Liver, and Blue-Green Algae Utilizing Randerath’s Method</td>
<td>66</td>
</tr>
<tr>
<td>3.</td>
<td>Nucleoside Composition of the T\textsubscript{1} RNase Generated Anticodon Containing Oligonucleotide</td>
<td>72</td>
</tr>
<tr>
<td>4.</td>
<td>Nucleoside Composition of Various T\textsubscript{1} RNase Generated Oligonucleotides</td>
<td>77</td>
</tr>
<tr>
<td>5.</td>
<td>Molar Ratios of Oligonucleotides of Euglena gracilis Cytoplasmic tRNAPhe</td>
<td>91</td>
</tr>
<tr>
<td>6.</td>
<td>5'-End Groups of Oligonucleotides of Euglena gracilis Cytoplasmic tRNAPhe</td>
<td>98</td>
</tr>
<tr>
<td>7.</td>
<td>Published M Values for One-Dimensional Electrophoresis on DEAE Paper at pH 3.5 or 1.9</td>
<td>102</td>
</tr>
<tr>
<td>8.</td>
<td>The pKa's and Average Charge (pH 3.5) of the 5'-Monophosphate Nucleotides</td>
<td>114</td>
</tr>
<tr>
<td>9.</td>
<td>The Sequence of T\textsubscript{1} and Pancreatic RNase Generated Oligonucleotides in the tRNAPhe of Euglena gracilis</td>
<td>158</td>
</tr>
<tr>
<td>10.</td>
<td>Positions in the tRNAs\textsuperscript{phe} Which Have Characteristic Eukaryotic Cytoplasmic or Prokaryotic Residues</td>
<td>175</td>
</tr>
<tr>
<td>11.</td>
<td>Sequence Homology Between Various Phenylalanine tRNAs\textsuperscript{phe}</td>
<td>180</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Generalized cloverleaf secondary structure of tRNA.</td>
<td>11</td>
</tr>
<tr>
<td>2.</td>
<td>Glycolytic reactions utilized in preparation of γ-[32P]-ATP</td>
<td>39</td>
</tr>
<tr>
<td>3.</td>
<td>Schematic representation of Randerath's tritium derivatization technique for base composition analysis</td>
<td>61</td>
</tr>
<tr>
<td>4.</td>
<td>Fluorographs of base composition analysis of the tRNAsPhe from (a) <em>Euglena gracilis</em> cytoplasm, (b) beef liver, (c) blue-green algae.</td>
<td>63</td>
</tr>
<tr>
<td>5.</td>
<td>Two-dimensional chromatograph of the T1 RNase digestion of <em>Euglena gracilis</em> cytoplasmic tRNAPhe</td>
<td>67</td>
</tr>
<tr>
<td>6.</td>
<td>Fluorographs of the base composition analysis of ACmUGmAAAYψCUAAGp.</td>
<td>69</td>
</tr>
<tr>
<td>7.</td>
<td>Fluorographs of the base composition analysis of DDG and Cm₂GψAG.</td>
<td>75</td>
</tr>
<tr>
<td>8.</td>
<td>Fluorograph of the base composition analysis of m₁AUm₅CCCG.</td>
<td>80</td>
</tr>
<tr>
<td>9.</td>
<td>Fluorographs of the sequence analysis by Randerath's tritium derivatization technique of m₁AUm₅CCCG.</td>
<td>83</td>
</tr>
<tr>
<td>10.</td>
<td>Autoradiographs of fingerprints obtained by two-dimensional electrophoresis of 5′-end [32P]-labelled oligonucleotides of <em>Euglena gracilis</em>.</td>
<td>88</td>
</tr>
<tr>
<td>11.</td>
<td>Standard maps of the mobilities of 5′-[32P]-labelled modified mononucleotides relative to the mobilities of standard uv marker mononucleotides.</td>
<td>93</td>
</tr>
<tr>
<td>12.</td>
<td>Autoradiographs of the 5′-end group analysis of the complete T1 RNase generated oligonucleotides of <em>Euglena gracilis</em> cytoplasmic tRNAPhe.</td>
<td>95</td>
</tr>
</tbody>
</table>
13. Autoradiograph of the sequence analysis by one-dimensional DEAE-paper electrophoresis at pH 3.5 of CACC and CACCA .................................................. 104

14. Autoradiograph of the sequence analysis by one-dimensional DEAE-paper electrophoresis at pH 1.9 of \( \psi \psi AG \) from Euglena cytoplasmic and beef liver tRNAs\( ^{Phe} \) ........................................ 107

15. Autoradiographs of the sequence analysis by one-dimensional DEAE-paper electrophoresis at pH 3.5 of CCG, AG, CUCAG, and \( \psi CG \) ........................................ 109

16. Effects of removal of nucleotides on electrophoretic mobility of parent oligonucleotides in two-dimensional homochromatography (angular mobility shifts); vertical differences between removal of purines and pyrimidines .......................... 114

17. Autoradiograph of one-dimensional homochromatography of partial SVP digestion of CUCAG ........................................ 117

18. Autoradiograph of two-dimensional homochromatography of the partial SVP digestion products of CCG ........................................ 121

19. Autoradiograph of two-dimensional homochromatography of the partial SVP digestion products of \( \psi CG \) ........................................ 123

20. Autoradiograph of two-dimensional homochromatography of the partial SVP digestion products of Cm\( _2 \psi AG \) ........................................ 125

21. Autoradiograph of two-dimensional homochromatography of the partial SVP digestion products of \( \psi \psi AG \) ........................................ 127

22. Autoradiograph of two-dimensional homochromatography of the partial SVP digestion products of CUCAG ........................................ 130

23. Autoradiograph of two-dimensional homochromatography of the partial SVP digestion products of T\( \psi CG \) ........................................ 132

24. Autoradiograph of two-dimensional homochromatography of the partial SVP digestion products of ACUUAm\( _2 \psi \) ........................................ 134

25. Autoradiograph of PEI-cellulose fingerprint of \( T_1 \) RNase generated oligonucleotides from Euglena gracilis ........................................ 137

26. Structure of the "peroxy" Y base of beef liver and wheat germ tRNAs\( ^{Phe} \) ........................................ 137
27. Autoradiograph of two-dimensional homochromatography of the combined partial SVP and partial Nuclease P<sub>1</sub> digestion products of AC<sup>m7</sup>U<sup>m</sup>G<sup>AAYA</sup><sub>ψ</sub>CUAAG on short plate (20 x 20 cm) .................................................. 140

28. Autoradiograph of two-dimensional homochromatography of the combined partial SVP and partial Nuclease P<sub>1</sub> digestion products of AC<sup>m7</sup>U<sup>m</sup>G<sup>AAYA</sup><sub>ψ</sub>CUAAG on long plate (20 x 40 cm) .................................................. 143

29. Autoradiograph of two-dimensional homochromatography of the combined partial SVP and partial Nuclease P<sub>1</sub> digestion products of G<sup>m</sup>AAYA<sub>ψ</sub> ................................................................. 146

30. Autoradiographs of two-dimensional homochromatography of the partial SVP digestion products of AAA<sup>m7</sup>G<sub>D</sub> from beef liver tRNA<sub>Phe</sub> before and after incubation in 50 mM ammonium bicarbonate, pH 9.0 at 85° C for 30 minutes .................................................. 149

31. Autoradiographs of the two-dimensional homochromatography of the partial SVP digestion products of m<sup>7</sup>GUCCCUG before and after incubation in 50 mM ammonium bicarbonate, pH 9.0 at 85° C for 30 minutes .................................................. 151

32. Autoradiography of the two-dimensional homochromatography of the partial SVP digestion products of m<sup>1</sup>AU<sup>m5</sup>CCCG before and after incubation in 50 mM ammonium bicarbonate, pH 9.0 at 85° C for 30 minutes .................................................. 154

33. Autoradiograph of the two-dimensional homochromatography of the combined partial SVP and Nuclease P<sub>1</sub> digestion products of GGGAGAG<sup>ψ</sup><sub>ψ</sub> ................................................................. 159

34. Autoradiograph of the two-dimensional homochromatography of the combined partial SVP and Nuclease P<sub>1</sub> digestion products of GCCGAGUU .................................................. 161

35. Autoradiograph of the two-dimensional homochromatography of the combined partial SVP and Nuclease P<sub>1</sub> digestion products of Am<sup>2</sup>GCUCAGDD .................................................. 163

36. Results of the sequence analysis of various oligonucleotides displayed so as to show overlaps and establishing the entire sequence of *Euglena gracilis* cytoplasmic tRNA<sub>Phe</sub> .................................................. 166

37. The nucleotide sequence of *Euglena gracilis* cytoplasmic tRNA<sub>Phe</sub> represented in cloverleaf form .................................................. 172
38. Comparison of the nucleotide sequences of
_Euglena gracilis_ cytoplasmic and chloroplastic
tRNAs^{Phe} represented in cloverleaf forms ........................... 172

39. Comparison of the nucleotide sequences of
(a) blue-green algae tRNAs^{Phe} and (b) bean
chloroplast tRNAs^{Phe} represented in cloverleaf
forms ................................................................. 178
## LIST OF ABBREVIATIONS

### Nucleosides

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>mA</td>
<td>1-methyladenosine</td>
</tr>
<tr>
<td>mA</td>
<td>6-methyladenosine</td>
</tr>
<tr>
<td>Cm</td>
<td>2'-O-methylcytidine</td>
</tr>
<tr>
<td>5C</td>
<td>5-methylcytidine</td>
</tr>
<tr>
<td>mG</td>
<td>N2-methylguanosine</td>
</tr>
<tr>
<td>2mG</td>
<td>N2,N2-dimethylguanosine</td>
</tr>
<tr>
<td>7G</td>
<td>7-methylguanosine</td>
</tr>
<tr>
<td>Gm</td>
<td>2'-O-methylguanosine</td>
</tr>
<tr>
<td>D</td>
<td>5,6-dihydrouridine</td>
</tr>
<tr>
<td>T</td>
<td>ribothymidine</td>
</tr>
<tr>
<td>ψ</td>
<td>pseudouridine</td>
</tr>
</tbody>
</table>

### Enzymes

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAP</td>
<td>bacterial alkaline phosphatase</td>
</tr>
<tr>
<td>SVP</td>
<td>snake venom phosphodiesterase</td>
</tr>
</tbody>
</table>

### Chemicals

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>PGA</td>
<td>3-phosphoglyceric acid</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>NTA</td>
<td>nitritotriacetic acid</td>
</tr>
<tr>
<td>TEA(B)</td>
<td>triethylamine (bicarbonate)</td>
</tr>
<tr>
<td>PPO</td>
<td>2,5-diphenyloxazole</td>
</tr>
</tbody>
</table>

**General**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>uv</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>cpm</td>
<td>counts per minute</td>
</tr>
<tr>
<td>TLC</td>
<td>thin layer chromatography</td>
</tr>
</tbody>
</table>
ABSTRACT

The establishment of the nucleotide sequence of the Euglena gracilis chloroplastic tRNA^phe by Chang et al. [Cell 9, 717 (1976)] was significant in that it was the first sequence established from an organelle. Since its sequence was generally more prokaryote-like than eukaryote-like, both in the lack of modified bases characteristic of eukaryotic tRNAs^phe and in the location of specific nucleotides within its sequence, this tRNA seemed to provide possible evidence for the endosymbiont theory of chloroplast origin. Any conclusions regarding the nature of chloroplast origin were considered mere speculation, however, without the sequence of the corresponding cytoplasmic tRNA^phe of Euglena gracilis.

The nucleotide sequence of Euglena gracilis cytoplasmic tRNA^phe was established mainly by utilizing various electrophoretic and chromatographic techniques to resolve oligonucleotides, which were labelled in vitro with $^{32}$P at their 5'-ends, and their partial digestion products. Oligonucleotides were generated by complete T$_1$ RNase digestions, and complete and partial pancreatic RNase digestions. Oligonucleotides were labelled using $\gamma$$-^{32}$P-ATP and polynucleotide kinase. Oligonucleotides were resolved utilizing two-dimensional electrophoresis, and analyzed for sequence utilizing partial digestion with snake venom phosphodiesterase or Nuclease P$_1$, followed by
two-dimensional homochromatography. Chemical cleavage of the tRNA at the phosphodiester bond adjacent to the 3'-end of its m^7G residue was utilized to create half-molecules whose analysis was useful in the ordering of smaller oligonucleotides. The 3'-end of intact *Euglena gracilis* cytoplasmic tRNA^phe^ was labelled by mild SVP treatment to remove part of its 3'-terminal CCA-OH end, and then labelling with [^32P] at the 3'-end using nucleotidyl transferase in the presence of α-[^32P]-ATP and non-radioactive CTP. Sequence analysis by partial digestion with Nuclease P_1 of the intact 3'-[^32P]-labelled tRNA, followed by two-dimensional homochromatography established the sequence of the first 13 nucleotides from the 3'-end.

The location of m^5C within the sequence of one T_(1) RNase generated oligonucleotide was established utilizing a chemical tritium derivatization technique developed by Randerath et al. A supplementary experiment was the base composition analysis of the total tRNA to confirm the presence of the amounts and identities of nucleotides expected from sequence analysis. In addition, various T_(1) RNase generated oligonucleotides were analyzed for base composition to verify the presence of modified bases. Base composition analysis was by Randerath's tritium derivatization technique.

The sequence of cytoplasmic tRNA^phe^ from *Euglena gracilis* is:

Both in the presence of characteristic modified nucleotides and the location of specific nucleotides within its sequence, the *Euglena* cytoplasmic tRNA\(^{\text{phe}}\) is a typical eukaryotic tRNA\(^{\text{phe}}\). Strikingly, it shares more sequence homology with tRNA\(^{\text{phe}}\) from mammalian sources than plant sources.

These findings have implications concerning the taxonomy of *Euglena gracilis*, the origin of the chloroplast, and the evolution of tRNAs.
INTRODUCTION

I. **Biological Functions of tRNA**

Even before the major biological function of transfer RNA had been identified, the need for an "adaptor molecule" in protein synthesis which both carries the amino acid as well as forms specific hydrogen bonds to the messenger RNA template was predicted by Crick (1). His hypothesis required one "adaptor molecule" for each amino acid, each adaptor having a specific hydrogen bonding surface which would enable it to bind specifically to the messenger RNA template, and one enzyme for each amino acid which would attach the amino acid to its specific "adaptor molecule." In 1958 Hoagland (2) identified the adaptor as transfer RNA (originally called soluble RNA because it was isolated from the soluble cytoplasmic fraction of cells), a class of small RNA molecules with molecular weights of about 25,000. The adaptor function of tRNA has been firmly established (3).

The typical cell type contains about 55 different species of tRNA (3) which are involved in protein synthesis. The need for more than 20 distinct tRNAs (the number of distinct amino acids which are coded for) reflects the degeneracy of the genetic code (4,5). Certain amino acids have multiple iso-accepting tRNA species thereby allowing for the translation of all possible codons for these amino acids.

As will be discussed below, the tRNAs as a group share
remarkable similarities at the primary, secondary, and probably
tertiary structural levels. These similarities are essential because
of the common molecular components which all tRNAs must interact with
as they perform their biological functions. Yet each tRNA species
must be structurally distinct enough to be specific for only one
amino acid. The specific amino-acylation ("charging") is carried out
by twenty different aminoaacyl-tRNA synthetases, each specific for one
amino acid and one set of iso-accepting tRNAs (6). These enzymes
require ATP for the initial activation of the amino acid before it
is transferred onto the 2' or 3' hydroxyl group (7,8) of the 3'
terminal adenosine.

During polypeptide synthesis the charged tRNA must interact
with a number of proteins. The methionine on the charged prokaryotic
initiator tRNA is formylated by the enzyme transformylase (3). The
charged initiator tRNA enters the ribosome as a ternary complex with
an initiation factor and GTP. It enters the initiation (I) site on
the 30S ribosomal subunit (which becomes part of the ribosomal P-site)
where it "decodes" the initiator triplet codon (3). Similarly, the
other charged tRNAs enter the ribosome in the form of a ternary
complex made with an elongation factor [EF-Tu in prokaryotes and EF1
in eukaryotes (6)] and GTP, and become located in the A-site of the
ribosome (3). On the ribosome, the tRNA interacts with various
proteins including peptidyl transferase (6), and is released after
having added its amino acid to the growing polypeptide chain of the
adjacent tRNA.
Transfer RNAs are involved in many other biological functions besides protein synthesis. When prokaryotic organisms are starved for amino acids, uncharged tRNAs become bound in the ribosomal A-sites setting off a signal for the formation of the unusual guanosine nucleotide derivatives ppGpp and pppGpp which serve as nucleotide messengers for the turning off of rRNA synthesis (9). Aminoacyl-tRNAs have been found in certain cases to play a role in the regulation of the biosynthesis of their specific amino acids. For example, in the Salmonella typhimurium histidine operon, charged tRNA\textsuperscript{his} acts as a co-repressor (10). A group of enzymes called aminoacyl-tRNA transferases have been shown to have the ability to catalyze the transfer of an amino acid from a charged tRNA to various acceptors. Acceptor molecules include the N-termini of intact proteins, phosphatidyl glycerol molecules (forming aminoacyl esters of phosphatidyl glycerol, a component of cell membranes), and N-acetyl muramyl peptides (intermediates in the synthesis of interpeptide bridges in bacterial cell walls) (11). Transfer RNAs have been found, in some cases, to function as allosteric effectors of enzyme activity. For example, a certain iso-acceptor of tRNA\textsuperscript{tyr} in Drosophila inhibits the enzyme tryptophan pyrrolase (12). Transfer RNA has been found to modify the activity of the E. coli enzyme endonuclease I from that of active double strand scission of DNA to low level single strand nicking (13). Eukaryotic tRNAs have been found to have some interesting properties with regard to viral metabolism. Reverse transcriptase from RNA tumor viruses has been found to use certain tRNAs as primers for the
synthesis of virally coded DNA (3). A number of selected tRNAs become incorporated (noncovalently) into RNA tumor virus particles during encapsulation from the cell membrane (3). Certain plant and animal viruses have been found to have "tRNA-like" 3'-ends which can actually be aminoaacylated by aminoaacyl-tRNA synthetases (6).

II. General Aspects of RNA Sequence Determination

Since the discovery of its central role in protein synthesis, transfer RNA has been the subject of intensive research aimed at elucidating its primary structure. Various techniques, ranging from the classical spectrophotometric to chromatographic and electrophoretic methods which utilize radioisotope technology, have been developed for the determination of nucleotide sequences. Yet all share a common basic approach. The purified molecule is first cleaved into oligonucleotide fragments of various lengths. These fragments are then resolved, and their molar ratios are determined. They are then sequenced by degradative (usually enzymatic) methods. Once the sequence of a sufficient number of fragments of various sizes has been established, the entire sequence can be deduced by analyzing overlaps. As a supplementary experiment, the whole molecule is degraded to the mononucleotide or nucleoside level and overall base composition is determined, as well as the identify and molar ratios for all modified bases.

Two base-specific ribonucleases are usually used for the initial cleavage of the total RNA molecule into fragments. These are
\( T_1 \) RNase (14) which cleaves at the 3'-end of guanine residues creating oligonucleotides which end in guanosine-3'-phosphates, and pancreatic RNase (15) which cleaves at the 3'-end of pyrimidines creating oligonucleotides which terminate in pyrimidine 3'-phosphates. By adjusting reaction conditions, one can obtain a complete \( T_1 \) RNase (14) or pancreatic RNase (15) digestion or a partial \( T_1 \) RNase (16) or pancreatic RNase digestion (17,18). (In the complete digestion all appropriate residues are cleaved, whereas in the partial digestion only the most accessible residues are cleaved.)

An additional method for generating long fragments is available for those tRNA molecules whose base composition analysis indicates the presence of \( \text{m}^7\text{G} \) (19). This residue, when present, always occupies a single unique position in loop III (see Fig. 1) and the specific chemical cleavage for \( \text{m}^7\text{G} \) described in Methods can be used to create half-molecules. These half-molecules can be sequenced individually as described above, and thus provide valuable information for the ordering of fragments in the whole molecule.

The most commonly used means of sequential degradation of oligonucleotide fragments is digestion with the 3'-exonuclease, snake venom phosphodiesterase (SVP) (20). (\( T_1 \) and pancreatic RNase generated oligonucleotides must first be treated with alkaline phosphatase, because SVP requires a free 3'-hydroxyl.) Nuclease \( P_{11} \), a relatively nonspecific endonuclease isolated from \( \text{Penicillium citrinum} \) (21), is useful for the digestion of oligonucleotides which contain modified residues which block SVP digestion. Randerath has developed a
chemical method for sequential degradation of oligonucleotides (22), which utilizes a phosphatase:periodate 3'-"pseudo-exonuclease" activity.

The fundamental differences between the various sequencing techniques arise in the methods for the detection and identification of the nucleotides as they are released during the sequential degradation of the oligonucleotide fragments. Other differences include the quantity of sample needed, and the methods for the isolation of oligonucleotide fragments.

The now classical approach to the sequencing of RNA involves the ultraviolet spectral identification (20) of nucleosides and mononucleotides. An advantage of this method is the absolute identification of a compound afforded by its ultraviolet spectrum. The great disadvantage of this method is that the sensitivity of detection is limited to the amount of compound necessary for ultraviolet identification. As much as 100-200 mg of a purified tRNA may be necessary to obtain a total sequence. The large amount of material required eliminates the possibility of sequencing tRNAs from many interesting sources which are available in only limited amounts. Furthermore, the fractionation of oligonucleotides and their enzymatic degradation produced at this large scale involves many column chromatography steps which are laborious and time consuming.

The development of electrophoretic and chromatographic techniques for the sequencing of in vivo uniformly $^{32}$P-labelled RNA by Sanger et al. (23,24) allowed for the rapid analysis of small
quantities of purified tRNA. Many tRNAs, particularly of prokaryotic origin, have been sequenced using these methods. RNAs from eukaryotic sources, however, are difficult to label \textit{in vivo} with $^{32}\text{P}$ to a specific activity needed for sequencing analysis. The solution to this problem has been the development of several methods for the \textit{in vitro} radio-chemical labelling of purified RNA samples.

Those \textit{in vitro} labelling (or post-labelling) techniques which utilize $^{32}\text{P}$ as the radiochemical agent use many of the electrophoretic and chromatographic methods in modified form originally developed for the \textit{in vivo} $[^{32}\text{P}]$-labelled RNA. These techniques are the 5'-end post-labelling of RNA with polynucleotide kinase (25,26) which was the primary technique used in the sequencing of both \textit{Euglena gracilis} chloroplastic and cytoplasmic tRNAs$^{\text{phe}}$ (27,28) and 3'-end post-labelling of RNA with polynucleotide phosphorylase (29).

The other major post-labelling technique is Randerath's novel tritium incorporation method (30,31,32) which was utilized for sequencing in this project to a minor extent, and was the method of labelling used for base composition analysis.

III. The General Structure of tRNA

The first structural information regarding tRNA came with the elucidation of the nucleotide sequence of yeast alanine tRNA by Holley (33) in 1965. Because physical studies of tRNAs indicated the presence of extensive intramolecular base pairing, Holley arranged the sequence so as to permit maximum hydrogen bonding, arriving at a
model for secondary structure which he termed the "cloverleaf" structure (see Fig. 1). To date, over 100 different tRNAs (specific for the various amino acids and isolated from diverse sources) have been sequenced (34), and every sequence determined can be accommodated within a "cloverleaf structure" (6). This information has provided strong circumstantial evidence for a common secondary structure for all tRNAs. Furthermore, sequence information has demonstrated other common structural features of tRNAs. In 1974, the first three-dimensional conformation of a tRNA, that of yeast tRNA\textsuperscript{Phe}, was deduced from 2.5 Å resolution X-ray diffraction analysis of two different crystalline forms (35,36,37). Although this three-dimensional model represents a static form of the tRNA molecule which functions dynamically \textit{in vivo}, enough information has been provided to consider structural features of the tRNA\textsuperscript{Phe} in functional terms. Furthermore, common structural features of tRNAs which have been gathered from sequence information suggest features of the three-dimensional structure of yeast tRNA\textsuperscript{Phe} which are common to all tRNAs.

Transfer RNA consists of a single polyribonucleotide chain ranging in size from 73 to 93 nucleotides (6). The cloverleaf structure is formed by the folding back upon itself of the polyribonucleotide chain with the formation of double helical stems (38) and single stranded loops. Together a stem and a loop are referred to as an arm. All tRNAs have four loops: dihydrouridine (D) loop (loop I), anticodon loop (loop II), variable loop (loop III), and the TψC loop (loop IV). All tRNAs contain at least four stems: the
acceptor stem, dihydrouridine stem, the anticodon stem, and the TΨC stem. Those tRNAs which contain a long variable arm have a fifth stem.

The acceptor stem usually contains seven base pairs with an additional four nucleotides (including the 3'-terminal CCA) protruding from one end. One important exception occurs in the acceptor stems of prokaryotic initiator tRNAs\textsuperscript{fmet} (6,39). These have been found to lack the Watson-Crick base pair at the end of the acceptor stem between the first nucleotide of the 5'-end to the fifth nucleotide from the 3'-end. The eukaryotic initiator tRNAs\textsuperscript{met} have the seventh base pair and it is always an A·U (6). It is interesting to note that the prokaryote \textit{Halobacterium cutirubrum} which, unlike most prokaryotic organisms does not use formylated methionine tRNA for the initiation of protein synthesis, uses an initiator tRNA\textsuperscript{met} which has a eukaryote-like stem (seven base pairs ending with A·U) (6,40).

The anticodon arm contains a stem made up of five base pairs, and a loop consisting of 7 nucleotides. These structural features seem to be essential for the proper alignment of the anticodon. Exceptions include tRNAs not involved in normal polypeptide synthesis. The tRNAs\textsuperscript{gly} from staphylococci (39) which are involved in cell wall biosynthesis and not protein synthesis have six base pairs in their anticodon stems. Those tRNAs which suppress frameshift mutations such as the Suf D mutant of \textit{Salmonella typhimurium} tRNA\textsuperscript{gly} have an extra nucleotide in their anticodon loop and recognize a quadruplet codon (41,42).
The TψC arm always contains a stem made up of five base pairs and a loop consisting of 7 nucleotides. The variability in tRNA size is primarily due to variations in the dihydrouridine arm and the variable arm. The dihydrouridine arm ranges in size from 15 to 18 nucleotides. Its stem contains three or four base pairs, and its loop contains from 7 to 11 nucleotides. There are two classes of variable arms: (1) those containing four or five bases which form a loop or (2) those which contain a large variable arm containing 13 to 21 nucleotides. Various classification schemes have been proposed for tRNAs based on the two variable regions of the cloverleaf structure (6,43,44).

Besides their generalized secondary structure, tRNAs share several invariant and semi-invariant (always a purine or pyrimidine) residues located in homologous positions in their sequences. In addition to the CCA at the 3' terminus, the following 12 nucleotides are invariant in almost all tRNAs involved in protein synthesis with the exception of the initiator tRNAs (the numbering system is that devised for yeast tRNA^phe): U_8, A_14, G_18, G_19, A_21, U_33, G_53, T_54, \psi_55, C_56, A_58, and C_61. Some of these invariant residues are found to be modified in certain species. For example, A_58 is often m^1A in eukaryotic tRNAs (39). There are eight semi-variant residues present in most tRNAs active in protein synthesis. These are Y_11, R_15, R_24, Y_32, R^*_37, Y_48, R_57, and Y_60 (where Y stands for pyrimidine, R stands for purine, and R^* stands for hypermodified purine). The positions of invariant and semi-invariant residues are illustrated in Fig. 1.
Fig. 1. Generalized cloverleaf structure of tRNA utilizing numbering system devised for phenylalanine tRNA (34). Circles represent nucleotides which are always present. Thick edged circles denote invariant or semi-invariant nucleotides (discussed in text). Ovals represent nucleotides which are not present in each tRNA sequence. Roman numerals refer to the four loops present in all tRNAs: loop I, the dihydrouridine loop; loop II, the anticodon loop; loop III, the variable loop; loop IV, the TψC loop.
The most notable exceptions to the occurrence of invariant and semi-invariant residues occur among tRNAs not involved in elongation of the polypeptide chain. The staphylococcal tRNAs\textsubscript{gly} which are not involved in protein synthesis at all lack some of the invariant residues (45). The G\textsubscript{18} and G\textsubscript{19} residues are replaced by U residues; the H\textsubscript{37} by either C or U and the $\psi$\textsubscript{55} by G. In some strains C\textsubscript{56} is replaced by U (6). Prokaryotic initiator tRNAs contain an A\textsubscript{11},U\textsubscript{24} base pair in the D stem rather than the semi-invariant Y\textsubscript{11},R\textsubscript{24}. Strikingly, the eukaryotic cytoplasmic initiator tRNAs lack the T\textsubscript{54}, $\psi$\textsubscript{55} [the G\textsubscript{53}, T\textsubscript{54}, $\psi$\textsubscript{55}, C\textsubscript{56} sequence found in most tRNAs has been shown to be involved in base pairing interactions with the ribosomes (46)], but instead have an A\textsubscript{54}, U\textsubscript{55} or U\textsuperscript{*}\textsubscript{55}. Also, eukaryotic cytoplasmic initiator tRNAs have an A\textsubscript{60} at the end of loop IV instead of the usual Y\textsubscript{60}. In some eukaryotic cytoplasmic initiator tRNAs the anticodon sequence CUA is preceded by C\textsubscript{33} instead of the usual U\textsubscript{33} (6).

The elucidation of the three dimensional structure of yeast tRNA\textsuperscript{phe} has shown how most of these invariant and semi-invariant residues are essential in maintaining the tertiary structure of tRNAs. For example, the two invariant G residues in loop I, G\textsubscript{18} and G\textsubscript{19}, have hydrogen bonding interactions with $\psi$\textsubscript{55} and C\textsubscript{56} respectively in loop IV (47).

Some general conclusions regarding tRNA structure were made by Rich and RajBhandary (6) using a synthesis of information from the three dimensional structure of yeast tRNA\textsuperscript{phe} and sequence data:
Extensive stacking interactions give the largest contribution to the stability of the molecule. These are primarily a result of tertiary hydrogen bonding interactions within loops and between loops involving both base-base interactions (usually not Watson-Crick base pairs) and base-backbone interactions. Stem regions consist largely of normal double helical RNA in the A conformation with only slight perturbations due to G·U pairs. The molecule is able to accommodate a certain amount of structural change such as alterations of hydrogen bonding between certain bases and still maintain its function (judging from sequence data). Variations in the D and variable arms are tolerable, but their biological roles are not known. In general, tRNAs are functionally designed. The anticodon bases are readily available for base pairing interactions with the m-RNA. The CCA 3'-terminus is flexible and single stranded so that it may be used in peptide bond formation. The interactions between loops I and IV (discussed above) are weak, and thus may allow the molecule to undergo conformational changes during protein synthesis.

IV. Structure-Function Relationships in tRNA

The great progress which has been made in elucidating tRNA structure is in sharp contrast with the presently scanty understanding of the precise role of structure in tRNA functions (48). For example, after a decade of intensive research, a precise understanding of the role of tRNA structure in the specific aminoacylation of each tRNA by its specific aa-tRNA synthetase has yet to be achieved. The
various aa-tRNA synthetases differ considerably from each other in size and subunit composition (49). The primary and tertiary structures of a few of these are presently being resolved (48). This knowledge coupled with what is known about tRNA structure would provide a basis for understanding the information which has been gathered by the research done to date. The original motivation of much of this research was to locate within the tRNA structure the aa-tRNA synthetase recognition site. The techniques used in locating this site included chemical modification studies, enzymatic dissections, sequence comparisons of iso-acceptors, studies of mutant tRNAs, and studies of aa-tRNA synthetase-tRNA complexes (48). The results seem to indicate that there is no such thing as a single recognition site. Goddard (48) hypothesizes that aminoacylation specificity involves more than one tRNA site and more than one recognition step. The initial interaction between tRNA and enzyme is non-specific and probably involves general symmetry recognition. It has been shown in vitro that tRNAs interact non-specifically with most synthetases, and may be misaminoacylated under certain conditions (50). Once initial interaction has been established, further interaction might proceed by, for example, specific hydrogen bonding. Slight conformational changes might occur in the tRNA and/or synthetase during progressive discriminating steps.

Similarly, knowledge of the role of tRNA structure in the mRNA-directed transfer of amino acids into proteins is still somewhat scanty. Nevertheless, due to ongoing research, certain details are
beginning to emerge. In prokaryotes, the structural basis for the interaction between the charged initiator tRNA$^{f_{\text{met}}}$ and IF2, the initiator factor, resides largely in the N-formyl methionine moiety (51). The charged initiator tRNA$^{f_{\text{met}}}$ does not accidentally complex with the elongation factor, EF-Tu because of one of its unique structural properties—it lacks a Watson-Crick base pair at the top of its acceptor stem. Schulman and co-workers demonstrated that when the tRNA$^{f_{\text{met}}}$ is chemically modified (Cl to Ul) to produce a base pair at the top of the acceptor stem, the aminoacylated modified tRNA$^{f_{\text{met}}}$ formed a ternary complex with EF-Tu and GTP (52). The conformation of tRNA is such as to allow the alignment of two tRNAs in the P and A sites so that their anticodons interact with adjacent codons on the messenger RNA while polypeptide and amino acid carrying 3'-ends are close enough to allow peptide bond formation (48). The interaction of anticodon and codon is specific for the first two positions of the codon. At the third position of the codon more than one base pair interaction is allowable because the base at the 5'-end of the anticodon is not as spatially confined as the other two, allowing it more freedom in forming base pairs with the 3'-end of the codon. The modified base found adjacent to the 3'-end of the anticodon probably stabilizes the conformation of the anticodon loop (53). Interaction between the tRNA and rRNA occurs by hydrogen bonding of the TΨC sequence of loop IV to the complementary sequence in the 5S rRNA of the 50S subunit. It is assumed that the required exposure of the TΨC occurs by a conformational change in the aminoacyl-tRNA on its
interaction with messenger RNA at the A-site of the ribosome (54).

The role of the relatively large proportion of modified bases found in tRNAs is still clouded in obscurity. It is known that when a modified base is present in the first position of the anticodon, pairing possibilities may be either increased or decreased. For example, inosine, I, may pair with U, C, or A, whereas 2-thiouridine can pair with only A (53) and not G. A role in stabilizing the anticodon loop has been described above for the hypermodified purine adjacent to the 3'-end of the anticodon. Although TψC is implicated in tRNA binding to the A-site by hydrogen bonding to the 5S rRNA, uridine modification does not appear to be essential (48). Feldman (55) has suggested that reversible (covalent) methylene crosslinks are formed from tRNA methyl groups which help fix the tRNA on the ribosome.

V. The Evolution of tRNA

The primary motivation for nucleotide sequence determination of tRNA was originally and is still largely an effort to understand structure-function relationships. As soon as the first few nucleotide sequences of tRNA were elucidated, it became apparent that a second type of information would emerge, that of the evolutionary relationships of tRNA. In 1966, Jukes (56) compared the sequences of the four tRNAs then known (all from yeast), those of alanine, tyrosine, and two serine tRNAs by alignment of homologous portions of their sequences with insertions of arbitrary gaps which inferred that
deletions of genetic material had taken place during evolution. Jukes hypothesized that the various tRNAs evolved from a common archetype by means of genetic duplication followed by functional differentiation. Homologous regions were preserved either due to the essential nature of their function, or because the time lapse following duplication had not been sufficient for the accumulation of enough randomly-occurring point mutations to produce differences in base composition.

In 1970, with the availability of the sequences of several E. coli tRNAs, McLaughlin (57) compared different tRNAs within the two major groups of organisms, prokaryotes and eukaryotes. He assumed that the time interval was the same for these comparisons because the divergence of the tRNAs took place in a common ancestor before the divergence of the eukaryotic and prokaryotic lines. He concluded that the overall rates of tRNA evolution in both lines were comparable.

By 1973, with over 40 tRNA sequences available for comparison, Holmquist and co-workers (58) were able to come to some general conclusions regarding the evolution of tRNA. Jukes' hypothesis of evolution by gene duplication and subsequent divergence seemed verified by similarities common to all tRNAs. Additional similarities were revealed by pairwise comparison of various tRNAs. These were greatest for certain pairs of iso-accepting tRNAs within the same species, but also occurred to an extent in certain pairs of tRNAs for different amino acids in the same organism. Furthermore, additional evidence for divergent evolution was shown with tRNAs for the same
amino acids in related organisms because these occasionally bear a sequence similarity that parallels the phylogenetic relationship of the organisms. For example, the valine tRNAs of the two strains of yeast, *S. cerevisiae* and *T. utilis* differ in only three nucleotide sites, whereas the phenylalanine tRNAs of *S. cerevisiae* and wheat germ, which are phylogenetically less similar, differ in 13 sites, and the two yeast (eukaryotic) valine tRNAs differ from the three *E. coli* (prokaryotic) valine tRNAs in about 33 sites. Although all tRNAs (with the exception of those involved in peptidoglycan synthesis) interact structurally with various components of polypeptide synthesis, there have been surprisingly few restrictions on the divergence of their primary structures. Holmquist (58) calculated that the average divergence for pairs of tRNAs for different amino acids involved in protein synthesis is 49.4%. This figure represents an equilibrium between selective and stochastic processes because according to his calculations, these tRNAs have had time to diverge to a maximum of about 75% (the divergence expected from stochastic processes alone is shown by comparing the two glycine tRNAs involved in peptidoglycan synthesis with tRNAs for different amino acids participating in polypeptide synthesis). Because certain structural features are necessary for tRNA function, natural selection has favored the more moderate divergence, and is responsible for the invariant and semi-invariant nucleotides in tRNAs (discussed above).

Comparison of nucleotide sequences has been used as a means of examining the phylogenetic relationships of the species whose tRNAs
have been sequenced (57, 58, 27). It has been difficult to construct coherent phylogenies based on this type of information alone because of the relatively wide divergence found in tRNAs as a class on the one hand, and because often a particular type of tRNA shows surprising conservatism among various species (27) on the other. One generalization seems to be valid, that is that there is a greater deal of sequence homology between eukaryotic tRNAs as a class and prokaryotic tRNAs as a class than between the two groups. Thus the sequences of tRNAs from fungi, for example, will show more homology with animal sequences than with bacterial sequences (57).

VI. Rationale for This Study

Biologists have come to realize that the most basic division among organisms is not that between animals and plants, but rather between eukaryotes (nucleated organisms) and prokaryotes (bacteria and blue-green algae). [Twenty years ago bacteria were still considered to be plants (59).] The most important distinction in the course of evolution, then, is the presence or absence of true nuclei, mitosis, meiosis, and certain cell organelles (57). Sequence studies of proteins and nucleic acids isolated from prokaryotes and eukaryotes have confirmed this major distinction (60), and have served as a means for the quantitation of this major divergence. McLaughlin (57), comparing sequences of various cytochrome c's and transfer RNAs, calculated that the divergence of prokaryotes and eukaryotes is 2.6 times more remote in evolution than the divergence of the eukaryotes into two separate kingdoms.
The discovery that two eukaryotic organelles, the mitochondria and chloroplasts, have their own DNA and have the functional apparatus for protein synthesis complete with ribosomes, tRNAs, and aminoacyl-tRNA synthetases has raised interesting questions regarding the evolutionary origin of organelles and its implications in the evolutionary divergence of eukaryotes and prokaryotes (61). It has been shown by various techniques that the DNA, ribosomes, tRNAs, and synthetases of these organelles differ from their cytoplasmic counterparts. It is known, however, that organelles are dependent on mutual interactions with the cytoplasm and nucleus of their "host" cells (62).

These observations have led many to believe that organelles evolved from prokaryotic endosymbionts (63). This view is not universally accepted, and another school of thought believes that organelles evolved through invagination and compartmentalization of function (64). One technique used to draw conclusions about the origins of organelles has been the comparison of nucleotide sequences of organelle RNAs and their cytoplasmic and prokaryotic counterparts. For example, homologies between the rRNAs of Euglena chloroplasts and those of several blue-green algae species were demonstrated by comparisons of sequences and by hybridization of algae rRNA to Euglena chloroplastic DNA (65). No homology was apparent between the rRNAs of Euglena chloroplasts and those of Euglena cytoplasm (66).

In 1976, we (27) reported the nucleotide sequence of Euglena gracilis chloroplastic tRNA^phe, the first primary sequence of an
organelle tRNA elucidated. We found that like all other phenylalanine tRNAs, *Euglena* chloroplast tRNA\(^{\text{phe}}\) contained 76 nucleotides. Furthermore, because it lacked the typical eukaryote tRNA\(^{\text{phe}}\) modified bases \(m^2G\), \(m^2G\), \(Cm\), \(Y\), and \(m^1A\), the chloroplast tRNA was more prokaryote-like than eukaryote-like. However, in other features of primary structure, such as the location of specific nucleotides within the sequence, *Euglena* chloroplast tRNA\(^{\text{phe}}\) had characteristics typical of both prokaryotic and eukaryotic tRNAs\(^{\text{phe}}\). The conclusion of this study was that, in general, *Euglena* chloroplast tRNA\(^{\text{phe}}\) is more homologous with prokaryotic tRNAs than with eukaryotic tRNAs. It should be noted that this conclusion was made by comparing the sequence of *Euglena* chloroplast tRNA\(^{\text{phe}}\) with only six other tRNAs\(^{\text{phe}}\) (three eukaryotic and three prokaryotic). Furthermore, the known tRNA\(^{\text{phe}}\) sequences as a class share a high degree of similarity. As far as dividing the tRNAs\(^{\text{phe}}\) into eukaryote-like and prokaryote-like groups, only six residues differ consistently between eukaryotes and prokaryotes. Of these six residues, *Euglena* chloroplast tRNA\(^{\text{phe}}\) resembles eukaryote tRNAs\(^{\text{phe}}\) in two positions (4 and 26) and prokaryote tRNAs\(^{\text{phe}}\) in four positions (20, 44, 45, and 60). Any conclusions made were considered speculative.

It was considered important to obtain more sequences of tRNA\(^{\text{phe}}\) for comparison in order to confirm any conclusions regarding the nature of the *Euglena* chloroplast tRNA\(^{\text{phe}}\). One sequence that was considered essential was the cytoplasmic tRNA\(^{\text{phe}}\) from *Euglena* which would provide a direct comparison of two tRNAs from the same organism,
one from the cytoplasm and one from the chloroplast. With this goal in mind, we embarked on the project described in this dissertation.
MATERIALS AND METHODS

I. Materials

A. Specialized Materials for Autoradiography, Chromatography, High Voltage Electrophoresis, and Radiation Safety

Kodak XR-5 x-ray film (20.3 x 25.4 cm and 35 x 43 cm) was purchased locally. 2,5-Diphenyloxazole (PPO) was purchased from New England Nuclear. Xylene cyanole FF dye was obtained from Eastman Kodak. Cellulose thin layer plates, glass or plastic backed, were from Analtech, Inc. Plastic backed 0.5% PEI-1000-cellulose sheets were prepared in the lab as described (67). PEI-1000 was kindly provided by Dr. K. Randerath. MN:500 cellulose (cellulosepulver) was purchased from Brinkman. Glass backed PEI-cellulose TLC plates were from Schleicher and Schuell, Inc. Glass backed DEAE-cellulose (Cell HR/DEAE; 15:2, 250 microns thick) TLC plates (20 x 20 and 20 x 40 cm) were purchased from Analtech, Inc.

Thin layer chromatography plates were all developed in glass tanks with the exception of the long DEAE-cellulose plates (20 x 40 cm) for homochromatography which were developed in a specially designed plexiglass tank manufactured by Wilbur Scientific Co., Boston, Mass.

Whatman (DE-81) DEAE paper, Whatman 3 MM paper, and Whatman #1 paper were bought from Reeve Angel and Company. Cellulose acetate strips (3 x 57 cm) were purchased from Schleicher and Schuell, Inc. Cellogel strips (3 x 57 cm) were purchased from Kalex, Inc. Cellogel
strips were stored in 30% methanol at 4° C and were soaked in 7 M urea, 2 mM EDTA, 5% pyridinium acetate, pH 3.5 for two hours before use.

DEAE-Sephadex A-25 was obtained from Pharmacia Fine Chemicals. Whatman DEAE (DE-32) cellulose was supplied by Reeve Angel and Company.

γ-[^32P]-ATP was prepared behind 2 inch thick plexiglass shielding sold and bent to specification by Gulf Wandas Plastics in Baton Rouge. A lead apron sold by Bar-Ray Products was worn during γ-[^32P]-ATP preparation.

B. Enzymes

Ribonuclease T₁ and T₂ were obtained from Sankyo Chemical Company. Both were dissolved in water (0.625 units/μl T₁ and 0.5 units/μl T₂) and stored at -20° C. Bovine pancreatic ribonuclease A was purchased from both Worthington Biochemical Corporation and Boehringer Mannheim Corporation. It was stored in a 0.25 μg/μl water solution at -20° C.

Snake venom phosphodiesterase was also bought from both Worthington and Boehringer. Boehringer enzyme was found unsuitable for use in the Randerath [³H] methods because it came stored in glycerol which interferes with the periodate oxidation of nucleases. Snake venom phosphodiesterase was generally stored as a 1 μg/μl water solution at -20° C.

Bacterial alkaline phosphatase and yeast hexokinase were
purchased as ammonium sulfate suspensions from Boehringer Mannheim Corporation, and could be stored undiluted at 4° C. Bacterial alkaline phosphatase was usually diluted to 2.5 units/ml with water and stored at -20° C. Hexokinase was diluted to 2.1 unit/ml in 50% glycerol, 5 mM sodium acetate, pH 5.0, and stored at -20° C.

Nuclease P₁ was purchased from Yamasa Shoyu Co. and stored in a 1.0 mg/ml 50 mM tris-maleate, pH 6.0 solution at -20° C. When ready for use, it was diluted to the proper concentration with 50 mM ammonium acetate, pH 5.3.

T₄ polynucleotide kinase was kindly provided by Dr. Simon Chang who purified it in the lab from an extract of E. coli strain B infected with T₄ phage amN82 according to the procedure of Richardson (68).

Ammonium sulfate suspensions of yeast phosphoglycerate kinase and rabbit muscle glyceraldehyde-3-phosphate dehydrogenase were purchased from Boehringer Mannheim Corporation and stored at 4° C.

C. Radioactive Materials and Fluors

Potassium boro[³H]hydride (15 Ci/mmole) was purchased from Amersham Corp. in lyophilized batches of 100 mCi. These were reconstituted with a solution of cold potassium borohydride (Sigma) in 0.1 N KOH giving a final concentration of 0.1 N potassium borohydride (5 Ci/mmole), 0.1 N KOH. This solution was then divided into 25 μl aliquots which were added to polyethylene tubes and lyophilized. These were stored at -80° C and reconstituted when needed.
Carrier free $[^{32}\text{P}]-\text{phosphate}$ in 20 mCi batches was bought from ICN Pharmaceuticals, Inc. or New England Nuclear and used in the preparation of $\gamma-[^{32}\text{P}]-\text{ATP}$ as described in Methods.

Carrier free sodium $[^{35}\text{S}]-\text{sulfate}$ for the making of radioactive ink was purchased from ICN Pharmaceuticals.

Aquasol and 2,5-Diphenyloxazole (PPO) were purchased from New England Nuclear.

D. Specialized Reagents and tRNA Samples

Nucleosides, potassium borohydride, and sodium periodate were purchased from Sigma Chemical Company.

Triethylamine was from Aldrich Chemical Company. It was made 2 M in water and then bubbled with CO$_2$ at 0°C until the pH was lowered to 7.5. Triethylamine bicarbonate (TEAB) was stored at 4°C in brown, opaque bottles.

Nucleoside 5' monophosphate and nucleoside 3',5' diphosphate uv markers were kindly provided by Dr. U. L. RajBhandary. Nucleoside triacetal uv markers were prepared as described in Methods.

Na$_2$-ATP was purchased from Sigma Chemical Company, dissolved in water, titrated to pH 7.0 with NaOH, adjusted to a final volume of 50 mM, and stored at -20°C.

Glutathionine (reduced form), 3-phosphoglyceric acid (PGA), and dithiothreitol (DTT) were from Sigma Chemical Company. Glutathione (40 mM), DTT (80 mM), and PGA (100 mM) were stored as water solutions at -20°C. Glutathione and DTT solutions were generally divided into
small aliquots stored in separate polyethylene tubes which after thawing and using once were discarded.

Nitrilotriacetic acid (NTA) was purchased from Schwarz Mann Chemical Company, dissolved in water, titrated to pH 7.0 with NaOH, adjusted to 50 mM final concentration, and stored at -20° C.

Aniline-HCl, pH 4.5 was prepared by titrating distilled aniline with concentrated HCl and was stored at a concentration of 0.3 M at -20° C.

Electrophoretic grade acrylamide and bisacrylamide were purchased from Bio-Rad Laboratories. Ultrapure urea was from Schwarz Mann.

Crude yeast RNA was purchased from Sigma, and stored at 4° C.

_Euglena gracilis_ chloroplast tRNA^{Phe} and cytoplasmic tRNA^{Phe} were isolated by Hecker and Barnett (69,70) at Oak Ridge National Laboratory, Tennessee. The latter tRNA was found to be insufficiently pure; therefore, it was further purified by RPC-5 chromatography (71).

E. Instrumentation and Equipment

Spectrophotometric determinations were made in a Gilford Model 200 spectrophotometer. High voltage electrophoresis was run in equipment from either Gilson Instruments, Inc. or Savant Instruments. Rotor vamping and evapomixing for γ-[32P]-ATP preparation were done utilizing a Flash Evaporator and a Evapo-Mix from Buchler Instruments, and a vacuum pump from General Electric.
Microquantities of liquids in polyethylene tubing were centrifuged with an Eppendorf Model 3200 microcentrifuge.

Liquid scintillation and Cherenkov counting were done with a Beckman LS-255 or LS-200 scintillation counter. A portable Geiger counter used for monitoring $^{32}$P activity and possible contamination was from Eberline Instrument Corporation.

II. Methods

A. General Methods

1. Autoradiography and Fluorography

Autoradiography of $^{32}$P samples on paper or TLC plates was performed using Kodak XR-5 x-ray film (20.3 x 25.4 cm or 35 x 43 cm) at room temperature in lead covered folders.

Length of exposure depended on amount of radioactivity and varied from 2 minutes for $\gamma$-$^{32}$P-ATP purification assays ($>1.0 \times 10^6$ cpm) to over a week for some two-dimensional homochromatography plates ($<2.0 \times 10^4$ cpm).

Fluorography of $^3$H samples on TLC plates was performed by first pouring a solution of 7% PPO in diethyl ether (w/v) over the surface of the plate and then air drying (72), and then wrapping film and plate in aluminum foil and exposing at $-80^\circ$ C. Exposure time was generally 3-4 days.

Radioactive ink was prepared by dissolving xylene cyanole FF dye in $^{35}$S sodium sulfate in water (0.2-2.0 mCi/ml).
2. Thin Layer Chromatography

Glass backed PEI-cellulose plates, to be developed in formic acid solvents, were prerun with 2 M pyridinium formate, pH 2.2, dried, prerun in water, dried, and then stored at 4° C (73). Plastic backed PEI-cellulose sheets, to be used for sequencing by [³H] labelling, were prepared as described (67). Cellulose plates and DEAE-cellulose plates were used directly from the box.

For cellulose plates, samples were generally applied in small aliquots of 0.5-1.0 μl in order to avoid diffusion. DEAE-cellulose and PEI-cellulose could accommodate larger aliquots (5.0-10.0 μl), if applied slowly, since diffusion is minimal on these ion exchange resins.

The following solvents were used in the various TLC systems described in subsequent sections:

(a) n-propanol:water:concentrated ammonium hydroxide (55:35:10, v/v), (prepared fresh daily)

(b) isobutyric acid:water:concentrated ammonium hydroxide (66:33:1, v/v), (prepared fresh daily)

(c) acetonitrile:4 N aqueous ammonia (3.4:1, v/v) (prepared fresh daily)

(d) t-amyl alcohol:methyl ethyl ketone:acetonitrile:ethyl acetate:water:formic acid, sp. grav. 1.2 (4:2:1.5:2:1.5:0.18, v/v), (prepared fresh daily)

(e) 0.8 M LiCl:0.8 M acetic acid (1:1, v/v), (stored at room temperature up to two months)

(f) 0.1 M potassium phosphate (pH 6.8):n-propanol:ammonium sulfate (100 ml:2 ml:60 grams), (0.1 M potassium phosphate, pH 6.8, stored at 4° C up to 6 months)
3. **Homochromatography**

i. **Preparation of the Homomix**

Homomixes of various strengths were prepared as described (74,75). For 0.5 liter, 210 grams of reagent grade urea were dissolved in 200 ml of deionized water at 37° C. Fifteen grams of crude yeast RNA were added and dissolved by stirring. The solution was brought to pH 7.0 utilizing 2 M KOH dropwise with stirring.

Additional 2 M KOH was added dropwise in an amount depending on final KOH concentration desired— from 10 mM to 75 mM KOH. (Final KOH concentration is dependent on the amount of KOH added after neutralization.) The mixture was covered with Saran wrap and incubated at 65° C for 25 hours. The solution was brought to 25° C and utilizing glacial acetic acid dropwise with constant stirring, was brought to pH 4.7. The volume was adjusted to 500 ml with water. Homomixes were stored at 4° C.

ii. **Development of DEAE-Cellulose Thin Layer Plates in Homomix**

All homochromatography (76) was performed in a 65° C oven in sealed TLC tanks with pre-heated solvents. Regular size DEAE-cellulose plates (20 x 20 cm) were run in glass tanks, while long DEAE-cellulose plates (20 x 40 cm) were run in specially designed plastic tanks (see Materials).

Whatman 3 MM paper (20 cm x 1/2 plate length) clamped to the top of the plate served as a wick. Plates were pre-heated in the oven and pre-run with water (about half way up the plate). They were then
transferred directly to the appropriate homomix solvent. The plates were run until the xylene cyanol dye (previously spotted at origin) reached 8.0 cm on the regular size plates and 36.0 cm on the long plates. The plates were then immediately oven dried. Visualization of homochromatography patterns was by autoradiography. Interpretation of patterns is explained in Results.

B. Methods for Base Composition and Sequence Analysis of RNA by Chemical Tritium Labelling

1. Preparation of Unlabelled Nucleoside Triacohol uv Markers

This technique was performed essentially as described (77). A 3.0 mM nucleoside solution (13.3 ml) was prepared consisting of equivalent molar amounts of uridine, cytidine, guanosine, and adenosine. Five μl of 2 M KOH were added (to get the purine nucleosides into solution).

A 200 μl aliquot of the nucleoside solution was oxidized with 5 μl 0.18 M NaIO₄ for 35 minutes at room temperature in the dark.

Then 5 μl 0.72 M NaBH₄ in 0.1 N NaOH were added. The mixture was incubated for 35 minutes at room temperature in the dark. Two hundred μl 1 N acetic acid were added to destroy NaBH₄. The solution was lyophilized and reconstituted in 100 μl water, and stored at -20° C.

2. Preparation and Isolation of Oligonucleotides by T₁ RNase Digestion

About 2.0 A₂₆₀ of tRNA were digested with 3.0 units of T₁ RNase in 0.02 M NH₄HCO₃, pH 7.8 (60 μl vol) for 7 hours at 37° C. Five μl bacterial alkaline phosphatase (2.5 units/ml) were then added, and the
mix was incubated for an additional 2 hours at 37° C. After extensive lyophilization to remove salt, the residue was reconstituted in 2 µl water. It was applied at 2½ cm from both the left and lower edges of a cellulose glass thin layer plate (20 x 20 cm).

The first dimension was run twice (79), first at 4° C in the cold room (12 hours), and then, after air drying, at room temperature (6 hours). The first dimension was run with solvent system (a).

After air drying, the second dimension (78) was run at room temperature (12 hours). The second dimension was run with solvent system (b).

Fragments were visualized under uv light. Extraction was by scraping spots, and suctioning the resulting powder into Eppendorf tips whose bottoms were plugged with glass wool. Sixty µl of water were added to the powder in each tip and then the tips were placed in the cold room for 2 hours. The solutions were then centrifuged into Eppendorf tubes. This step was repeated once. The solutions were measured for absorbance at 260 nm in a Gilford spectrophotometer.

3. Base Composition Analysis of Ribopolynucleotides

This technique was performed essentially as described by Randerath (31,77,79) with four minor modifications.

a) Polyethylene tubes (1.5 ml) were substituted for glass test tubes for running reactions as well as for storage of potassium boro [3H] hydride. Recovery of sample was much better in polyethylene tubes, and they were also more convenient for lyophilization.

b) Samples were incubated with 0.35 units of T2-RNase for 2 hours at 37° C without buffer before digestion with
Randerath's enzyme mix (77). Samples treated this way gave more accurate results because the T₃ RNase digests certain linkages more efficiently than the enzyme mix.

c. Reduction of volume for TLC was by lyophilization rather than by blowing with N₂.

d. Instead of a 2-fold molar excess of periodate over nucleoside and a 5-fold molar excess of borohydride over periodate, a 3-fold excess of periodate over nucleoside and a 20-fold excess of borohydride over periodate were used (80).

Generally, 0.1 nmole of ribopolynucleotide (equivalent to about 0.1 A₂₆₀ of intact tRNA) was digested to the nucleoside level for base analysis. The RNA (in 10 μl water) was first incubated with 0.35 unit of T₃ RNase for 2 hours, 37° C. The solution was then lyophilized and reconstituted with 5 μl of an enzyme-buffer mix (77) consisting of 0.064 μg/μl snake venom phosphodiesterase, 0.064 μg/μl pancreatic ribonuclease A, 0.048 μg/μl bacterial alkaline phosphatase, 3 x 10⁻² M bicine-Na (the buffer added was 0.6 M, pH 8.0), 1 x 10⁻² M MgCl₂, and incubated for 12 hours, 37° C.

The resulting nucleosides were oxidized with the addition of an approximate 3-fold molar excess of NaIO₄ over nucleoside (20 μl total volume). Incubation was at room temperature for 2 hours in the dark. The incubation mixture was then cooled on ice and 1.5 μl 0.15 M phosphate, pH 6.8 was added. A 20-fold excess of potassium boro [³H] hydride was then immediately added. (The potassium boro [³H] hydride used was 0.1 N in 0.1 N KOH, and had a specific activity of 5 Ci/mmole.) Incubation was in the dark for 2 hours at room temperature. Excess borohydride was destroyed by the addition of 2 μl 5 N acetic acid.
This operation was performed in the hood as tritium gas was given off. The solution was allowed to sit for 20 minutes, and then was lyophilized. The residue of labelled trialcohols was taken up in 2 \( \mu l \) 0.1 N formic acid and stored at -20° C.

In preparation for two-dimensional cellulose TLC (81), 4 \( \mu l \) water were added to digests of tRNA whereas 4 \( \mu l \) cold trialcohol uv markers, prepared as described above, were added to digests of fragments.

Aliquots for TLC (1 \( \mu l \) for tRNA digests; 2 \( \mu l \) for fragments) were applied at 2.5 cm each from the left hand and bottom edge of a plastic-backed cellulose TLC plate (20 x 20 cm). The first dimension was run with solvent system (c) (81), 17 cm from the origin. The TLC plate was air dried and a Whatman \#1 wick was attached to the top of the sheet by stapling. The second dimension was run with solvent system (d) (81), and development was to 3 cm on the wick.

After air drying of plates, wicks were removed and markers were visualized under uv light and outlined to serve as reference spots when comparing to the standard map (81).

Ten ml of 7% PPO in diethyl ether were poured over the plate to serve as fluor for autoradiography. Exposure to film was for two to three days. Compounds were located by aligning the plate with the film and cut out.

Elutions of cutouts were performed layer side down in 25 ml beakers with 1 ml 2 N ammonia for 3 hours. Then 0.5 ml aliquots were transferred to scintillation vials to which 10 ml of aquasol were
added. Samples were quantitated by liquid scintillation counting.

4. Sequence Analysis of Oligonucleotides by Partial Snake Venom Phosphodiesterase Digestion and ³H Labelling

Tritium sequence analysis of partial SVP digests (32) proved useful for the unambiguous location of any modified residues not found at the 5'-end of an oligonucleotide.

The partial digest mixture contained 0.005 mM oligonucleotide, 10 mM magnesium acetate, 20 mM bicine, pH 8.0, 0.1 μg/μl bacterial alkaline phosphatase, and 0.01 μg/μl SVP in a 100 μl volume. Digestion was at room temperature. Five μl of 0.75 mM NaIO₄ were added to 10 μl aliquots of the digestion taken at 0, 0.5, 1, 2, 3, 5, 10, 20, 25, and 35 minutes. Oxidation was at room temperature in the dark for 1 hour. One μl 0.1 M potassium boro [³H] hydride (5 Ci/mmol) was added to each aliquot. The reduction mixes were incubated for 1 hour at room temperature in the dark. Fifteen μl 5 N acetic acid were added to each aliquot in the hood (tritium gas was given off). After 30 minutes, the solutions were lyophilized, and then reconstituted in 10 μl water. Samples were stored at -80° C until analysis.

Analysis of labelled solutions was by PEI-cellulose TLC (22) of aliquots corresponding to 0.02 nmole of original oligonucleotide. Preparation of plastic backed PEI-cellulose plates (20 x 35 cm) was as described (22,67).

Prior to application of samples, the PEI-cellulose sheet was shaked in 300 ml methanol/conc. ammonia (1000:1, by vol.) for 5 minutes and air dried. After application, the sheet was pre-developed
with water to 3-4 cm on a Whatman #1 wick attached at 22 cm above the origin. Development was in Randerath's 8.0 M urea, 0.05-0.8 M tris-HCl, pH 8.0 gradient solvent system (67). After development, the sheet was air dried and then soaked with agitation in 500 ml methanol. Visualization was by fluorography following coating of plates with 7% PPO in diethyl ether (10 ml). Release of the labelled 3'-termini nucleoside trialcohols was accomplished by in situ digestion with RNase T2 (2-4 μl, 1mg/ml) at 25° C for 12 hours. The enzyme was applied directly to the spot, which was then covered with a piece of Parafilm. A heavy glass plate was placed on top of this assemblage to prevent evaporation.

The labelled trialcohol nucleosides were extracted from the plate by scraping each spot, and suctioning the resulting powder into an Eppendorf tip with glass wool at its bottom. One hundred and twenty μl of water were added to the powder and it was allowed to sit in the cold room for 2 hours. The tip was then placed into an Eppendorf tube and the liquid was centrifuged into the tube. The solution was then lyophilized and reconstituted in 4 μl uv markers (see above), 1 μl 0.1 N formic acid, and was stored at -20° C. Identification of the released [3H] labelled trialcohol nucleosides was by using the base composition analysis TLC system described above and then comparing to a standard map (81).

C. Methods for Sequencing Using 5'-End-Group Labelling with [32P]

1. Preparation of γ-[32P]-ATP
Commercial $\gamma-^{32}\text{P}}\text{-ATP}$, purchased from ICN Pharmaceuticals, Inc. was found to be unsuitable for our use due to a contaminant which inhibited the polynucleotide kinase reaction. $\gamma-^{32}\text{P}}\text{-ATP}$ was, therefore, prepared in our lab enzymatically utilizing carrier-free $^{32}\text{P}}\text{-inorganic phosphate}$ and an exchange reaction reported by Glynn and Chappell (82). This involved the glycolytic reactions shown in Fig. 2.

As indicated by the $\Delta G^\circ$'s, the equilibrium position of the combined reaction is far to the left. Utilizing the conditions detailed below, the reaction can initially be forced to the right. In the presence of excess $^{32}\text{P}}\text{-phosphoric acid}$ a large proportion of the $\gamma$-phosphate positions of ATP exchange for the radioactive phosphate as equilibrium is established.

A mixture (200 $\mu$l vol) was prepared containing 0.25 M tris-HCl (the buffer added was 1.0 M, pH 7.5), 35 mM MgCl$_2$, 2 mM ATP (added as 50 mM, pH 7.0), 10 mM glutathione, and 2.5 mM 3-phosphoglyceric acid. An enzyme mixture (15 $\mu$l vol) containing 6.6 $\mu$g glyceraldehydepophosphate dehydrogenase, and 3.3 $\mu$g phosphoglycerate kinase was prepared. Carrier free $^{32}\text{P}}\text{-phosphoric acid}$ (20 mCi in 50 $\mu$l) was transferred from the shipping container to the reaction tube, a 15 ml polystyrene test tube. Fifteen $\mu$l of the cocktail were added to the tube. The exchange reaction was initiated by the addition of 3 $\mu$l of the enzyme mixture. The reaction was allowed to proceed at room temperature (usually 20 minutes).

In order to monitor the extent of incorporation of $^{32}\text{P}}$ into
Fig. 2. Glycolytic reactions utilized in preparation of $\gamma$-$^{32}$P-ATP. Reaction A is catalyzed by phosphoglycerate kinase, and reaction B is catalyzed by glyceraldehyde dehydrogenase. As can be seen from the $\Delta G^\circ$ (83) of the combined reaction, equilibrium strongly favors the reverse reaction. Conditions are detailed in the text whereby the reaction can initially be forced to the right. As equilibrium is re-established excess $^{32}$P-phosphoric acid present in the cocktail is incorporated into the $\gamma$-phosphate position of ATP by simple exchange.
40

\[ \text{3-Phosphoglycerate} + \text{ATP} \rightleftharpoons \text{3-Phosphoglyceroyl phosphate} \]
\[ \Delta G^\circ = 4.5 \text{ kcal mol}^{-1} \]

\[ \text{3-Phosphoglyceroyl phosphate} + \text{NADH} + \text{H}^+ \rightleftharpoons \text{Glyceraldehyde 3-phosphate} + \text{NAD}^+ + \text{P}_i \]
\[ \Delta G^\circ = -1.5 \text{ kcal mol}^{-1} \]

Combined Reaction:
\[ \text{3-Phosphoglycerate} + \text{ATP} + \text{NADH} + \text{H}^+ \rightleftharpoons \text{Glyceraldehyde 3-phosphate} + \text{P}_i + \text{ADP} + \text{NAD}^+ \]
\[ \Delta G^\circ = 3.0 \text{ kcal mol}^{-1} \]
ATP, 1 μl time aliquots were taken from the reaction tube at zero

time (before addition of enzyme mix), 2, 5, 10, and 20 minutes after

addition of enzyme mixture and pipetted into separate tubes containing

1.9 ml of a Norite charcoal solution. The Norite solution consisted

of 10% Norite A (acid washed), 25 mM HCl, 0.5 mM sodium phosphate, and

0.5 mM sodium pyrophosphate. The Norite tubes were then vortexed and
centrifuged and 1 μl aliquots were taken from the supernatant,
applied to filter paper squares, and Cherenkov counted. Because of
its aromatic moiety, ATP is strongly absorbed by the charcoal, whereas
free phosphate has little absorption (the added cold phosphate has the
effect of swamping out any binding of [32P]). Therefore, as the
reaction proceeds, the counts in the supernatant go from a high at
zero time to a lower figure at equilibrium. The percent incorporation
can, thus, be calculated. Typical yields ranged from 55-82%. The
reaction was stopped by diluting to 0.5 ml with water.

The reaction mixture was then loaded onto a 4 ml bed volume
DEAE-Sephadex A25 column previously equilibrated with 0.05 M TEAB,
pH 7.5. A 150 ml gradient of 0.05 M to 1.0 M TEAB, pH 7.5 was applied
to the column. A fraction collector was adjusted to collect
1.5-2.0 ml/tube (usually 3-6 min./tube).

One μl aliquots were taken from each tube, applied to filter
paper squares, and Cherenkov counted. The results were plotted as
tube number vs. cpm. Before the ATP tubes were combined, one μl
aliquots from each were sometimes taken and spotted on a PEI TLC
plate and run in solvent (e) to check for purity. Radioactive spots
were visualized by autoradiography. ATP tube contents were combined, transferred to a round bottom flask and repeatedly evaporated with the addition of water at 35° C in a Buchler Flash Evaporator in order to get rid of the TEAB. This process was continued until TEAB salt was no longer visible as the solution dried.

The solution was transferred to a conical centrifuge tube, evaporated and reconstituted several times in a Buchler Evapomix Instrument at 35° C, neutralized with a few µl of 0.1 M NaOH, brought to a final volume of 0.5 or 1.0 ml made 30% ethanol and 5.0 mM tris with a final pH of about 7.0.

Specific activities varied from about 0.6 x 10^6 to 1.2 x 10^6 cpm/pmole. Purity checks using the PEI TLC system described above were run to check for phosphate contamination of the ATP solution. γ-[³²P]-ATP solution was stable for up to a month at -20° C and longer at -80° C.

2. Chemical Cleavage of the Phosphodiester Bond at the 3'-End of m^7G

When base composition analysis of a tRNA species indicated the presence of m^7G (which when present always occupies a single position in loop III in all tRNA's sequenced to date), this technique (19,25) was used to create half-molecules which were useful for sequence deduction.

The conditions detailed below allowed for the specific cleavage of a m^7G-containing tRNA at the 3' phosphodiester bond of this residue. Alkaline incubation opened the m^7G imidazole ring and
made the ribosyl bond sensitive to acid hydrolysis. Acid addition depurinated the residue. Aniline incubation catalyzed the cleavage of the 3' phosphodiester bond by β-elimination.

Before chemical cleavage, 5.0-6.0 A260 units of tRNA were dialyzed against 10⁻⁴ M EDTA to remove salt. The tRNA was then incubated in 50 mM NaOH (pH 12) in a volume of 200 μl for 15 minutes at room temperature. The addition of 3 μl 5 N acetic acid brought the pH to 4.5. Two hundred μl 0.3 N aniline-HCl (pH 4.5) were then added, and the mixture was incubated for 4 hours at 37° C. Reactions were stopped by freezing in dry ice. Samples could at this point be stored at -20° C.

In preparation for separation of cleavage products by polyacrylamide gel electrophoresis, samples were desalted on small DEAE-cellulose columns. These were prepared by packing acid washed Carlsburg disposable pipettes (4 cm packing length). Columns were equilibrated with 50 mM TEAB, pH 7.5-8.5.

The samples were diluted to 5 ml with water, and loaded onto the columns at a flow rate of 0.5 ml/minute. The columns were run with 50 mM TEAB until no Cl⁻ could be detected in the effluent with 10% HNO₃/5% AgNO₃. Samples were eluted from the column with 2 M TEAB, pH 7.5-8.5 at a flow rate of 0.1-0.2 ml/minute; fractions of about 0.8 ml each were collected by hand and read at 260 nm in a Gilford spectrophotometer. Appropriate fractions were combined and lyophilized several times to remove TEAB. Separation of the cleavage products was by polyacrylamide gel electrophoresis.
3. **Separation of $m^7G$ Cleavage Products by Polyacrylamide Gel Electrophoresis**

Polyacrylamide gel slabs, 15% acrylamide (95% recrystallized acrylamide, 5% N, N'-methyl-bis-acrylamide), were prepared and run as described by Akrod (84) with the modification that polymerization was in the presence of 7 M urea which kept denatured samples from renaturing (85). Mobility was, therefore, a strict function of chain length rather than conformation.

Stock solutions of the buffer system used in making the slabs and running the electrophoresis consisted of 108 g of tris, 0.3 g of disodium-EDTA and 55 g of boric acid in 1.0 liter, pH 8.3 (adjusted with HCl) (86).

The 15% gel slab was prepared as follows. A mixture was made consisting of 35.7 g ultrapure urea, 8.5 ml buffer, pH 8.3 (described above), 12.1 g recrystallized acrylamide, 0.6 g bis-acrylamide and enough water to bring the total volume to 85 ml. Eighty-five µl of N, N, N', N'-tetramethylethylenediamine (accelerator) and 30 µl 50% ammonium persulfate (catalyst) were added to the mixture. After three minutes, the mixture was poured into a gel mold (20 x 20 cm) (84) containing slot templates (1 cm wide each). Under these conditions, the gel solidified in about 30 minutes. The gel was prerun for 30 minutes at 10 mA, 4° C in a ten-fold dilution of the stock buffer to remove impurities and polymerization catalysts.

Cleavage samples were lyophilized to dryness, and reconstituted in 25 µl of loading solution consisting of 7 M urea,
40% sucrose, 0.1% bromphenol blue, and 0.1% xylene cyanole FF all in a ten-fold dilution of the stock buffer.

Samples were denatured by boiling at 100° C, 1 minute, and quickly chilling on dry ice.

For preparative slots, 12 μl of sample prepared as above were applied. For analytical slots (for staining), 1.0 μl of sample plus 10 μl of loading solution were applied.

Intact tRNA, 0.1 A₂₆₀ unit in 12 μl loading solution, were run in an adjacent slot to serve as a marker and to determine qualitatively the extent of the cleavage reaction.

The gel was run 6 hours at 380 v (7.5 mA), then 11 hours at 500 v (12 mA), or until the bromphenol blue was at the end of the gel.

Analytical and intact tRNA slots were cut away from preparative slots, fixed in 1.0 N acetic acid for ½ hour, then stained in 0.2% toluidine blue in 15% v/v acetic acid. Destaining of excess dye was in a circulating bath of 0.2 N acetic acid for 12 hours.

Samples (0.1 A₂₆₀ unit minimum) in preparative slots were visualized by ultraviolet light after having first aligned the preparative slots next to the analytical slots to determine the approximate location of sample bands.

The method of elution of cleavage products from gel pieces (87) consisted of grinding excised gel in 2 ml of 0.3 M NaCl, 30 mM sodium citrate, pH 8.0 in a motor driven homogenizer with teflon pestle for 3 minutes on ice, centrifugation of the resulting suspension,
and saving of the supernatant. This process was repeated and the two supernatants were pooled. The supernatants were then diluted and desalted on small DEAE-cellulose columns as described above.

More recently, a less tedious technique has been suggested for elution of sample from gel bands (75, 88) which consists of electrophoresis of the sample directly from the excised gel band into a dialysis bag.

4. Partial Pancreatic RNase Digestion of tRNA

This technique (17, 18) was performed to obtain long oligonucleotide fragments whose sequencing would provide valuable overlap information for ordering fragments obtained by complete digestions. One A$_{260}$ unit of tRNA was digested partially with 50 ng of pancreatic RNase at 0°C for one hour in 30 mM tris, pH 8.0. Pancreatic RNase was then removed by repeated extraction with phenol. The phenol phase was twice back extracted with 0.1 M tris, pH 8.0. Phenol was removed by extraction with diethyl ether (3 X). The tRNA was precipitated with two volumes of absolute ethanol at -20°C.

The partial pancreatic digestion was reconstituted in water. About 0.1 A$_{260}$ of digestion material was incubated with 1.25 x 10$^{-3}$ units of bacterial alkaline phosphatase in 50 mM tris, pH 8.0 for 2 hours at 37°C. Two μl of 0.05 M nitrilotriacetic acid (NTA) were added, and the solution was allowed to sit at room temperature for 20 minutes. The solution was then boiled at 100°C for ninety seconds to inactivate the phosphatase.
Oligonucleotides thus generated were labelled at their 5'‐ends with \[^{32}P\] (as described in section 6). Separations of the 5'‐end group \[^{32}P\]‐labelled oligonucleotides was by the two‐dimensional polyacrylamide gel electrophoresis system described by Wachter and Fiers (89). This system utilizes a 10% acrylamide acid gel (25 mM citric acid, pH 3.5, 6 M urea, 4 mM EDTA) in the first dimension, and a 20% acrylamide neutral gel (40 mM Tris‐citric acid, pH 8.0) in the second dimension. Fragments were visualized by autoradiography, eluted by grinding excised gel in 1 M NaCl, and then desalted in small DEAE cellulose columns as described in the previous section.

5. Complete Digestion of tRNA by T₁ or Pancreatic RNase

The techniques described in this and the following section were originated by Sanger's group (90) and modified to their present form by Dr. U. L. RajBhandary and co‐workers (26,75). For T₁ RNase digestion, 0.05 A_{260} units of purified tRNA were incubated in 15 µl of a 35 mM Tris, pH 8.0 solution containing 0.625 units of T₁ RNase for 3 hours at 37° C. At this time, 2.5 x 10⁻³ units of bacterial alkaline phosphatase (BAP) were added to 0.5 µl, and the digest was incubated for an additional 2 hours at 37° C. The digest was then brought to 25° C and 2 µl of 0.05 M nitrilotriacetic acid (NTA), a chelating agent for zinc, were added. After 20 minutes, the solution was subjected to 100° C for ninety seconds, and then rapidly frozen in crushed dry ice. NTA treatment was necessary for inactivation of the phosphatase because the apoenzyme form is more susceptible to heat denaturation.
In the case of pancreatic RNase digestion the above procedure is repeated in its entirety with 0.25 μg of pancreatic RNase replacing the T₁ RNase.

The T₁ and pancreatic RNase digestions were stored at -20° C until needed.

6. [³²P]-Labelling of the 5'-'Ends of T₁ or Pancreatic RNase Generated Oligonucleotide Fragments by Polynucleotide Kinase

The 5'-ends of RNase digestion fragments were labelled with [³²P] using polynucleotide kinase (68, 91) and γ-[³²P]-ATP.

Into a 1.5 ml Eppendorf tube, 1 x 10⁹ cpm of γ-[³²P]-ATP were transferred. To this was added enough cold ATP to bring the total pmoles of ATP to 1700. Equal volume of water was added to dilute the ethanol and thus prevent splattering during lyophilization. The solution was frozen in crushed dry ice and lyophilized.

To the lyophilized ATP was added 1 μl 80 mM dithiothreitol, 1 μl 0.1 MgCl₂, 5 μl water, 2 μl of RNA digest (0.005 A₂₆₀ unit), and 0.5 μl of 7 unit/μl polynucleotide kinase. The mixture was incubated at 37° C for 30 minutes. Then, 2.5 μl of 10 mM glucose and 1 μl 2 U/ml hexokinase were added to destroy γ-[³²P]-ATP (25, 92). This mixture was incubated for another 20 minutes at 37° C with two chases of 1 μl 2.5 mM ATP at five minute intervals starting after addition of glucose and hexokinase. The mixture was then brought to 100° C for 60 seconds to denature enzymes and stored at -80° C.
7. Two-Dimensional Resolution of 5'-End $^{32}$P-Labelled Fragments—Fingerprinting

RNase digestion fragments labelled at their 5'-ends with $^{32}$P as described above were separated in a high voltage two-dimensional electrophoresis system essentially identical to that introduced by Sanger et al. (73) with minor modifications introduced by RajBhandary's group (73,75). For analytical scale runs, 1 µl of labelled material was used directly. For larger scale preparative runs the entire labelling reaction mixture was lyophilized to dryness and reconstituted in 1.5 µl water.

The first dimension was run on a cellulose acetate strip (3 x 57 cm) previously soaked in a 2 mM EDTA, 7 M urea, 5% pyridinium acetate, pH 3.5 buffer. The sample was run in a pre-heated (30-35° C) tank from 5 to 23 cm on the strip at 5,000 volts (about one hour), as measured by the progress of the blue component of xylene cyanol tracking dye spotted adjacent to the sample.

The second dimension was run on a sheet of DEAE-cellulose (DE-81) paper (46 x 57 cm) saturated with 7% formic acid at 120 mA from 6 to 26 cm also measured by following the blue dye (12-14 hours). (The pH of 7% formic acid in the tank was 1.9.)

The transfer of sample from the first to the second dimension was accomplished by laying the cellulose acetate strip face down flush with a width-wise line 6 cm from one end of the DEAE-81 sheet which in turn sat on a stack of 4 Whatman 3-MM strips (22.5 x 2.5 in.) centered about the 6 cm line. Another stack of
4 Whatman 3-MM strips (22.5 x 1.75 in.) previously water saturated was laid on top of the cellulose acetate strip. A glass sheet was placed on the entire assembly, and pressure was applied by hand. The glass was periodically lifted, more water applied to the top stack of strips, the glass lowered and pressure reapplied.

The DEAE sheet was then aligned on a rack and the transfer region was washed with a gentle stream of absolute ethanol to flush away urea, and allowed to dry.

Saturation of the DEAE sheet with 7% formic acid was accomplished by placing a plexiglass rod under the sheet along the transfer line and two more rods over the sheet on either side of the transfer line, and gently spraying the sheet with 7% formic acid on the outer sides of the two top rods. This was continued until the sample transfer line was saturated with buffer. Direct spraying of buffer onto the sample area would smear the sample. The paper was then aligned on the rack and the rest is sprayed with 7% formic acid.

When better resolution was desired, cellogel strips were substituted for cellulose acetate strips. These have a higher capacity for sample and have been recommended for samples containing more than 0.2 M salt and/or more than 0.2 A_260 units of RNA (73).

Alternatively, a thin layer system substituted for the DEAE paper in the second dimension was sometimes used to increase resolution. In this case the first dimension was transferred to a polyethyleneimine cellulose plate which was run in 2 M pyridinium formate, pH 3.4 as described (93). Whatman 3-MM paper was clamped
to the top of the TLC plate to serve as a wick. Xylene cyanol tracking dye was allowed to travel 12 cm from the origin.

Both the above procedures produced more compact spots, and were employed for separating fragments which otherwise only partially resolved or streaked. The latter procedure had the added advantage of speed (about 6 hours). These techniques could be combined, cellogel in the first dimension, PEI TLC in the second.

Fragments were visualized by autoradiography, cut out or scraped, and quantitated by Cherenkov counting.

Fragments were eluted from DEAE paper according to Barrell's procedure (24) using 2 M TEAB directly into 1.5 ml Eppendorf tubes.

Fragments were scraped from PEI TLC plates, suctioned into Eppendorf pipette tips plugged with glass wool, and allowed to sit in 10 μl 2 M TEAB for 4 hours at 4° C. Another 10 μl TEAB were then added, and the eluate was centrifuged into Eppendorf tubes from the tips. Recovery was greater than 95%.

Crude yeast carrier RNA (0.5 A_260 unit) was added to each fragment to minimize the loss of radioactive sample by absorption to the sides of the tubes, and, also to slow down the partial digestions described in another section.

TEAB was removed by repeated lyophilization. Eluted fragments were stored at -20° C at an initial concentration of 5,000 cpm/μl, if feasible.
8. *5'-End Group Determination of [$^{32}$P]-Labelled Fragments*

Digestion of fragments to the mononucleotide level utilizing T$_2$ RNase, snake venom phosphodiesterase (SVP), or P$_1$ nuclease (26,73, 94) followed by thin layer chromatography were utilized in identifying the radioactively labelled 5' ends.

For T$_2$ RNase digestions, 5,000-10,000 cpm of each [$^{32}$P]-labelled fragment were incubated with 0.5 unit of T$_2$ RNase in 30 mM ammonium acetate, pH 4.6 for two hours at 37° C. The digest was then lyophilized to dryness and reconstituted in a 2 µl uv marker solution containing 0.1 A$_{260}$ unit each of pAp, pUp, and pCp. Two cellulose TLC plates were then spotted with 1 µl of each fragment. One plate was run with solvent system (a). The other plate was run with solvent system (f). Plates were run until the solvent front reached 1 cm from the top of each plate (usually 6-7 hours for either system). The plates were air dried and the standard markers were located under uv light and circled. Radioactive spots were located by autoradiography.

For the snake venom phosphodiesterase digestion, 5,000-10,000 cpm of each [$^{32}$P]-labelled fragment were incubated with 2 µg SVP/0.1 A$_{260}$ unit of yeast carrier RNA in a 50 mM tris-HCl, 5 mM potassium phosphate, pH 8.9 buffer at 37° C for 2 hours. The digests were then lyophilized to dryness and reconstituted in a 2 µl uv marker solution containing 0.1 A$_{260}$ unit each of pA, pU, pC, and pG. The SVP digests were then analyzed in the same chromatographic systems described above for the T$_2$ RNase digests.
For \( P_1 \) nuclease digestions, which create 5'-mononucleotide end groups identical to SVP digestions, 5,000-10,000 cmp of \( ^{32}P \)-labelled fragment were incubated with 0.02 \( \mu \)g \( P_1 \) nuclease/\( A_{260} \) unit of yeast carrier RNA in a 50 mM ammonium acetate, pH 5.3 buffer at 37° C for 2 hours. These digests were co-chromatographed with the same marker system described for SVP digests above.

Assignments of 5'-ends were made by comparing the results with standard maps (see Results). Occasionally, to obtain better resolution, digests reconstituted in marker solution were subjected to two-dimensional TLC utilizing the same two solvent systems mentioned above with the isobutyric system in the first dimension and the ammonium sulfate system in the second.

9. Partial Digestion of 5'-End \( ^{32}P \)-Labelled Fragments with Snake Venom Phosphodiesterase or \( P_1 \) Nuclease

Reaction conditions described above for complete digestions with SVP or \( P_1 \) nuclease were modified by using less enzyme and by running the reactions at room temperature in order to slow the reactions down so that time aliquots could be taken among which all successive stages of digestion were contained.

For partial SVP digestions, labelled oligonucleotide solutions were made 50 mM tris-HCl, 5 mM potassium phosphate, pH 8.9, and 1.0 \( \mu \)g SVP/0.1 \( A_{260} \) unit of yeast carrier RNA was added.

A minimum of 5,000 Cherenkov cpm/time aliquot were desired for analysis. Typically, 10 time aliquots were taken from 0 (before SVP addition) to 160 minutes. Therefore, a minimum of 50,000
Cherenkov cpm total were desirable for each partial digest. Ideally, 100,000-200,000 Cherenkov cpm of labelled fragment were digested in a total volume of 50 µl for 160 minutes with 5 µl aliquots taken at 0, 1, 2, 5, 10, 20, 40, 60, 80, and 160 minutes. If less than 50,000 cpm were available for a particular fragment, the total volume was reduced appropriately and fewer aliquots were taken (time intervals were chosen during which all digestion products were guessed to be present). If the total counts were so low that 5,000 cpm/chosen aliquot were not available, autoradiography of the DEAE-cellulose TLC plates used for the analysis of partial digests (described below) was made more sensitive by pouring a solution of 7% 2,5 diphenyl-oxazole (PPO) in diethyl ether over the plates before exposing the films. Using this technique we were able to sequence a fragment (see pg.156) with less than 5,000 cpm total.

As aliquots were taken, they were added to individual Eppendorf tubes containing 1 µl 0.01 M EDTA, boiled at 100°C for 2 minutes (to inactivate SVP), and frozen in crushed dry ice.

For partial P₁ nuclease digestion, labelled oligonucleotide solutions were made 50 mM ammonium acetate, pH 5.3, and 0.01 µg P₁ nuclease/A₂₆₀ unit of yeast carrier RNA was added.

P₁ nuclease functions as a random endonuclease; the whole range of digestion products was usually available within minutes of enzyme addition. Typically, 5 µl aliquots were taken at 0, 2, 5, 10, and 20 minutes. The incubation mixture usually had a total volume of 25 µl and contained 30,000 Cherenkov cpm and up of labelled fragment.
Time aliquots were transferred to individual Eppendorf tubes containing 1 µl 0.01 M EDTA, boiled at 100° C for 4 minutes, and frozen in dry ice.

Both SVP and P₁ digests were stored at -20° C until analysis.

10. Sequence Analysis of 5'-End [³²P]-Labelled Partial Digests

Two techniques were available for the sequence analysis of [³²P]-labelled partial digestion products. Generally, fragments were analyzed both ways.

i. One-Dimensional High Voltage Electrophoresis

This technique was run as described by Barrell (24) on DEAE-cellulose (DE-81) paper in either the 5% pyridinium acetate, pH 3.5 tank or the 7% formic acid, pH 1.9 tank depending on the resolution requirements of the particular fragment. DEAE-cellulose has the effect of binding SVP to its matrix, and thus, effectively immobilizes this enzyme. Therefore, it was not necessary to boil time aliquots to inactivate the enzyme. Time aliquots were spotted directly onto the paper as they were taken. Samples were spotted along the 6 cm line along with xylene cyanol tracking dye, and 5'-end uv markers (pA, pU, pC, and pG, 1.0-2.0 A₂₆₀ units of mixture/spot). Paper was wetted with appropriate buffer solution for electrophoresis as described above for fingerprinting. Electrophoresis was run at 100 mA until the blue dye reached 1.5 inches before midpoint of the sheet (10-15 hours). The sheet was dried in the hood; patterns were visualized by autoradiography.
Often, partial digestions of fragments of similar established sequence were run alongside for comparison.

Sequence assignments were made on the basis of characteristic M values described in Results.

ii. Two-Dimensional Homochromatography

This technique, originally introduced by Sanger et al. (95) for DNA sequencing, was modified for RNA sequencing by RajBhandary and coworkers (73,75).

Before actual sequence analysis, the extent of digestion was tested by one-dimensional homochromatography (76) of 1.0 µl samples of each time aliquot followed by autoradiography. The appropriate time aliquots as well as quantities needed could then be chosen to be pooled for two-dimensional homochromatography so that approximately equal amounts of each digestion intermediate could be displayed. This procedure was sometimes omitted when the total Cherenkov cpm of a particular fragment were low (see previous section).

In preparation for two-dimensional homochromatography, pooled time aliquots were lyophilized to dryness, and reconstituted in 1.5 µl water.

Conditions for the first dimension were essentially identical to those described for fingerprinting, except that sample was applied in 5 or 6 small spots across the width at 15 cm on the cellulose acetate strip instead of a single spot to avoid streaking due to salt, and that the sample was run from 15 to 25 cm as
measured by the blue component of the xylene cyanol tracking dye. This process took about 30 minutes.

The second dimension was run on DEAE-cellulose TLC plates by homochromatography as described above. Regular size DEAE-cellulose plates (20 x 20 cm) were found to be suitable for fragments up to 12 nucleotides long. Longer fragments were analyzed on long DEAE-cellulose plates (20 x 40 cm).

The transfer of sample from first to second dimension was done as follows. The cellulose acetate strip was laid lengthwise directly over a 1 ml pipette taped to the bench. This had the effect of elevating the midline of the strip. In order to wash the sample (which after electrophoresis is present in a broad belt along the strip) onto the midline of the strip, a stack of three Whatman 3-MM strips (23 x 4.5 cm), previously water saturated, was laid along and slightly over either (lengthwise) edge. Water was gently sprayed onto the Whatman 3-MM stacks along the edges not touching the cellulose acetate strip. The TLC plate was then placed face down onto the pipette-cellulose acetate strip assemblage with the midline of the strip in contact with a width-wise line on the plate 2 cm from one end. The plate was supported at the other end by another pipette. A weight was put onto the plate. This assemblage was allowed to sit for about 30 minutes. After drying, the plate was ready for homochromatography.

Interpretation of two-dimensional homochromatography patterns is discussed in Results.
RESULTS

I. Base Composition Analysis of *Euglena gracilis* Cytoplasmic tRNA^phe by Chemical Tritium Labelling

Hecker and co-workers (69) compared base compositions of chloroplast and cytoplasmic tRNA^phe's from *Euglena gracilis*. The compositions were determined after hydrolysis to free nucleosides utilizing phosphatase treatment of alkaline hydrolysates or combined action of T_1 RNase followed by snake venom phosphodiesterase and phosphatase. The nucleosides were resolved on a uniform-particle-size-cation-exchange resin (96) at pH 4.6. Nucleosides were identified by their spectral ratios as well as positions of elution (97).

Their results (see Table 1) were utilized as a guide in the sequence analysis of these tRNA's. Once a tentative sequence of *Euglena* cytoplasmic tRNA^phe had been assigned, it was decided for several reasons to re-examine its base composition utilizing a different technique. The number of guanosine and cytidine residues in Hecker's determination varied slightly from what we expected (see Table 1) from our sequence assignment. Also, dihydouridine could not be identified by the spectrophotometric assay. Furthermore, we were interested in comparing Hecker's results with those obtained utilizing different experimental conditions.

We chose Randerath's tritium derivatization technique (31,77) as our method of base composition analysis. This technique has the
Table 1. Comparison of the Nucleoside Compositions of tRNAPhe from the Chloroplasts and Cytoplasm of Euglena gracilis Determined Spectrophotometrically by Hecker (69) and Nucleoside Composition Expected for Cytoplasmic tRNAPhe from Sequence Assignment

<table>
<thead>
<tr>
<th>Nucleoside</th>
<th>tRNAChl Hecker</th>
<th>tRNA Cyt Hecker</th>
<th>tRNA Cyt expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>19</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>U</td>
<td>13</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>C</td>
<td>18</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td>G</td>
<td>21</td>
<td>21</td>
<td>19</td>
</tr>
<tr>
<td>mA</td>
<td>0</td>
<td>1.1</td>
<td>1</td>
</tr>
<tr>
<td>ψ</td>
<td>1.3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>T</td>
<td>1.1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Da</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>5C</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>m2G</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>m2G</td>
<td>0</td>
<td>1.2</td>
<td>1</td>
</tr>
<tr>
<td>m7G</td>
<td>0.5</td>
<td>0.3</td>
<td>1</td>
</tr>
<tr>
<td>Y</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ms216A</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cm</td>
<td>0.9</td>
<td>0.9</td>
<td>1</td>
</tr>
<tr>
<td>Cm</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Am</td>
<td>0</td>
<td>0.9</td>
<td>0</td>
</tr>
<tr>
<td>X</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*aDihydouridine cannot be determined spectrophotometrically*
advantages of great sensitivity (as little as 1.0 μg of polyribo-
nucleotide required), high-resolving power of nucleoside derivatives,
and the accurate and precise quantitation afforded by radioactive
labelling.

The technique consists of the following steps:

1. enzymatic digestion of the ribopoly nucleotide to a
mixture of nucleosides
2. oxidation of the nucleoside mixture with periodate
to nucleoside dialdehydes
3. reduction with tritiated borohydride to labelled
nucleoside trialcohols
4. two-dimensional cellulose thin-layer chromatography
5. visualization of tritium labelled nucleoside
derivatives by fluorography
6. elution of labelled nucleoside derivatives and
quantitation by liquid scintillation counting

Fig. 3 is a schematic representation of these steps showing
the structures of the nucleoside dialdehydes and the [\(^3\)H]-labelled
trialcohols.

The experimental details of Randerath's tritium derivative
base composition technique, as well as some minor modifications we
added, have been described in Materials and Methods.

Fig. 4 shows fluorographs of tRNA\(^{\text{phe}}\) base composition analyses
from a) Euglena gracilis cytoplasm, b) beef liver, and c) blue-green
algae. These were obtained as described in Materials and Methods, and
the compounds were qualitatively identified by comparing their
positions on the fluorographs to a standard map (81). The apostrophe
Fig. 3. Schematic representation of Randerath's (31) tritium derivatization technique for base composition analysis.
Ribopolynucleotide

1) T2-RNase
   37°C, 2 hours

2) Ribonuclease A
   Snake Venom Phosphodiesterase
   Alkaline Phosphatase
   pH 8.0, 37°C, 12 hours

Nucleosides

NaIO₄
pH 6, 22°C, 2 hours

Nucleoside Dialdehydes

[³⁴H]-KBH₄
pH 7-8, 22°C, 2 hours

[³H]-labelled Trialkohols

Two-Dimensional Thin Layer Chromatography, Fluorography, Elution, and Scintillation Counting
Fig. 4. Fluorographs of base composition analyses by Randerath's (31) tritium derivatization technique of the tRNAPheS from (a) *Euglena gracilis* cytoplasm, (b) beef liver, and (b) blue-green algae.
symbol, "", refers to the tritiated nucleoside derivatives of the indicated compounds. $\psi$-$D$ is a decomposition compound related to $\psi'$ (81). When computing the moles of $\psi$, the cpm of $\psi'$-$D$ and $\psi'$ are combined.

Beef liver tRNA was analyzed to serve as a comparison with *Euglena* cyt. tRNA$^{\text{phe}}$ because of its similar sequence (see Fig. 38). Conversely, blue-green algae tRNA$^{\text{phe}}$, a typical prokaryotic tRNA lacking the modified bases $m^2G$, $m^2G$, $Y$, and $m^1A$, differs considerably from these eukaryotic tRNA's.

Table 2 contains the nucleoside compositions which were calculated using the radioactivities of the compounds extracted from these thin layer cellulose chromatograms compared with the base compositions expected from sequence analyses. Low yields of dihydrouridine in the *Euglena* and beef liver analyses were probably due to overdosing conditions of borohydride (98). Generally, correlation of the observed to the expected base compositions in these experiments matched or exceeded that of similar published results (101,102).

II. Base Composition Analysis of T$_1$-RNase Fragments

T$_1$ RNase digestion fragments were prepared and isolated from 2.0 A$_{260}$ units of *Euglena gracilis* cytoplasmic tRNA$^{\text{phe}}$ as described in section B-2 of Methods. Fig. 5 is a tracing of the two-dimensional chromatogram of this digest in which the encircled areas represent the various uv absorbing spots.
Table 2. Nucleoside Composition\(^a\) of tRNA\(^\text{phe}\)'s of *Euglena* Cytoplasm, Beef Liver, and Blue-Green Algae Utilizing Randerath's Method

<table>
<thead>
<tr>
<th>Nucleoside</th>
<th><em>Euglena</em> Cytoplasm</th>
<th>Beef Liver</th>
<th>Blue-Green Algae</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>17</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>U</td>
<td>8.2</td>
<td>8</td>
<td>10.7</td>
</tr>
<tr>
<td>C</td>
<td>16.2</td>
<td>17</td>
<td>15.7</td>
</tr>
<tr>
<td>G</td>
<td>18.6</td>
<td>19</td>
<td>18.0</td>
</tr>
<tr>
<td>(^1\text{m}A)</td>
<td>1.1</td>
<td>1</td>
<td>1.9</td>
</tr>
<tr>
<td>(\psi)</td>
<td>4.0</td>
<td>5</td>
<td>3.4</td>
</tr>
<tr>
<td>T</td>
<td>0.8</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>D</td>
<td>1.2</td>
<td>2</td>
<td>1.9</td>
</tr>
<tr>
<td>(^5\text{m}C)</td>
<td>1.1</td>
<td>1</td>
<td>1.4</td>
</tr>
<tr>
<td>(^2\text{m}G)</td>
<td>0.8</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>(^2\text{m}G)</td>
<td>1.0</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>(^7\text{m}G)</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Y</td>
<td>0.4</td>
<td>1</td>
<td>0.4</td>
</tr>
</tbody>
</table>

\(^a\)For the cytoplasm tRNA\(^\text{phe}\) relative abundance was calculated assuming one mole of \(^2\text{m}G\) per tRNA. Similarly, for beef liver and blue-green algae tRNA\(^\text{phe}\)'s, the expected numbers of G and U respectively were utilized in calculating the relative abundance.

\(^b\)Thirty percent of beef liver tRNA\(^\text{phe}\) has a \(\psi\text{CG}\) instead of \(T\psi\text{CG}\) (99,100).
Fig. 5. Tracing of two-dimensional chromatograph of the T₁ RNase digest of *Euglena gracilis* cytoplasmic tRNAPhe where encircled spots represent various uv absorbing fragments. Dashed circle shows location of xylene cyanol tracer dye.
Spot A was thought to contain the 15 nucleotide long anticodon containing T₁ RNase generated fragment (which we had previously assigned a tentative sequence by 5'-end[^32P]-labelling techniques) for two reasons. First, its position on the chromatogram indicated that it was the slowest moving fragment in both dimensions as would be expected for the longest T₁ fragment in the molecule. Secondly, under uv light it gave off a fluorescent sheen unlike the other fragments, indicating the presence of the fluorescent Y base. Experimental evidence detailed in subsequent sections indicated that this fragment had an identical sequence to that of the T₁ anticodon fragment of beef liver tRNAₚhe, ACᵐUGᵐAAYAψCUAAAGp. A weakness of the[^32P]-labelling techniques utilized in sequencing this fragment is that there is no direct identification of modified bases if they do not occur at the 5'-end of the oligonucleotide. In this fragment the existence of Gᵐ at the given position can be inferred because the T₁ digestion did not give a ACᵐUGp. Similarly, the existence of Cᵐ at the given position is verified by the fact that the pancreatic RNase digestion did not produce an oligonucleotide which was cleaved at the 3'-end of this residue. No similar inference can be made for ψ which is cleaved as efficiently by pancreatic RNase as other pyrimidines. Furthermore, the ψ jump is indistinguishable from the U jump in the two-dimensional homochromatography system used for analyzing 5'-end[^32P]-labelled fragments. Y base does have a characteristic jump (103), however, additional confirmatory evidence is desirable.

A base composition analysis of spot A utilizing Randerath's
Tritium derivatization technique was, therefore, performed to confirm the presence of ψ and Y. (This method cannot confirm the presence of G\textsuperscript{m} or C\textsuperscript{m} since 2'-O-methylated nucleosides are not susceptible to periodate oxidation.) In the original experiment, 0.1 nmole of spot A (about 0.015 A\textsubscript{260} unit) was digested solely with Randerath's enzyme mixture (31) consisting of snake venom phosphodiesterase, pancreatic RNase, and bacterial alkaline phosphatase. The results of the base composition analysis, performed as described in Materials and Methods, were qualitatively correct, demonstrating the presence of both ψ and Y. The molar ratios, calculated assuming one G per oligonucleotide, were not, however, as expected (see Table 3-a). This was thought to be due to incomplete digestion to the nucleoside level because the snake venom phosphodiesterase is hindered by the presence of certain modified bases including ψ (104,105) and Y (see pg. 136). In order to obtain better molar ratios, we decided to try a preliminary T\textsubscript{2} RNase digest before the addition of the usual enzyme-buffer mixture (31). T\textsubscript{2} RNase digests are usually run in ammonium acetate buffer, pH 4.6. However, even small traces of ammonia interfere with periodate oxidation (106). We, therefore, digested two aliquots of spot A (0.1 nmole each) with T\textsubscript{2} RNase—one in 6 mM ammonium acetate, pH 4.6, and one without buffer. Digestions were for 2 hours at 37° C. The digest with buffer was then subjected to extensive lyophilization. Randerath's enzyme-buffer mixture was then added to each digest. Conditions for this and all subsequent steps were identical to those described in Materials and Methods.
Table 3 gives the base compositions calculated for spot A
a) without preliminary T₂ RNase digestion, b) with T₂ RNase digestion
and buffer, c) with T₂ RNase digestion and no buffer. Table 3 shows
that although all three experiments provided qualitative evidence for
the presence of ψ and Y, the ratios closest to those expected were
obtained when a preliminary T₂ RNase digestion without buffer was run.
Preliminary T₂ RNase digestion has, therefore, become standard
procedure in our lab for base composition analysis. Fig. 6 shows the
fluorograph of the base composition analysis of spot A.

Another fragment which needed verification of sequence
assignment by base composition analysis was the T₁ RNase generated
fragment, DDGp. Other eukaryotic tRNA^phe's including yeast, wheat
germin, and mammalian contain this sequence, whereas the prokaryote
and Euglena chloroplast tRNA^phe's contain the T₁ RNase generated
fragment, DUGp. [³²P]-labelling techniques allowed for the
unambiguous identification of the 5'-end as D. However, these
techniques were incapable of distinguishing between DDGp and DUGp.
Base composition analysis of spot I confirmed the sequence DDGp as
can be seen by referring to Fig. 7a and Table 4.

An additional fragment which needed verification of sequence
assignment by base composition analysis was the T₁ RNase generated
fragment, ψAGp or occasionally Cm₂GψAGp. (Some T₁ RNase digestions
we performed cleaved at the 3'-end of m₂G, whereas others did not.
Fingerprints either contained a Cm₂G spot and a ψAG spot or a single
Cm₂GψAG spot. Never was a fingerprint obtained which contained all
Table 3. Comparison of Nucleoside Composition of Spot A Obtained with a) no preliminary T2 RNase digestion, b) preliminary T2 RNase digestion in 6 mM ammonium acetate, pH 4.6, c) preliminary T2 RNase digestion with no buffer

<table>
<thead>
<tr>
<th>Nucleoside</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3.8</td>
<td>13.2</td>
<td>6.6</td>
<td>7</td>
</tr>
<tr>
<td>C</td>
<td>1.4</td>
<td>3.5</td>
<td>1.7</td>
<td>1</td>
</tr>
<tr>
<td>U</td>
<td>1.8</td>
<td>4.5</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>G</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>\psi</td>
<td>0.3</td>
<td>1.0</td>
<td>0.6</td>
<td>1</td>
</tr>
<tr>
<td>Y</td>
<td>0.3</td>
<td>1.3</td>
<td>0.4</td>
<td>1</td>
</tr>
<tr>
<td>G\text{m}</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>C\text{m}</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
</tbody>
</table>
Fig. 6. Fluorograph of the base composition analysis by Randerath's tritium derivatization technique (31) of spot A, the T₁ RNase generated anticodon containing oligonucleotide.
Fig. 7. Fluorographs of the base composition analysis by Randerath's tritium derivatization technique (31) of (a) spot I, DDG and (b) spot D, Cm\textsubscript{2}G\psi\textsubscript{AG} with contaminants (see text).
Table 4. Nucleoside Composition of Spots I<sup>a</sup>, D<sup>b</sup> and G<sup>c</sup>

<table>
<thead>
<tr>
<th>Nucleoside</th>
<th>Spot I</th>
<th>Spot D</th>
<th>Spot G</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-</td>
<td>-</td>
<td>2.2</td>
</tr>
<tr>
<td>C</td>
<td>-</td>
<td>-</td>
<td>3.1</td>
</tr>
<tr>
<td>U</td>
<td>-</td>
<td>-</td>
<td>1.2</td>
</tr>
<tr>
<td>G</td>
<td>1.0</td>
<td>1</td>
<td>3.4</td>
</tr>
<tr>
<td>ψ</td>
<td>-</td>
<td>-</td>
<td>3.0</td>
</tr>
<tr>
<td>D</td>
<td>1.9</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>m&lt;sub&gt;2&lt;/sub&gt;C</td>
<td>-</td>
<td>-</td>
<td>0.9</td>
</tr>
<tr>
<td>m&lt;sub&gt;1&lt;/sub&gt;A</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>m&lt;sub&gt;5&lt;/sub&gt;C</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<sup>a</sup>Spot I is DDG.

<sup>b</sup>Spot D contains Cm<sub>2</sub>GψAG, CUCAG, and ψCG (see text).

<sup>c</sup>Spot G contains m<sub>1</sub>AUm<sub>5</sub>CCCCG
three spots. Although we never investigated this phenomenon, we assumed that under properly controlled conditions, T₁ RNase will cleave at m₂G.) In those fingerprints in which the Cm₂G occurs separately, it is easily distinguishable from CG by its position on the fingerprint. Furthermore, base composition analysis of the total cytoplasmic tRNA^phe showed the presence of one mole of m₂G per mole of tRNA. The problem arose in proving that the tetranucleotide was actually ψψAGp rather than ψUAGp. We suspected the former sequence for a number of reasons. Mammalian tRNA^phe contains the T₁ fragment, ψψAGp, and there is no example of ψUAGp in any tRNA^phe yet sequenced.

Under complete SVP digestion conditions for the 5'-end group analysis of the [³²P]-labelled fragment, two spots showed on the chromatogram (see Fig. 12, digest 11b). This indicated that even under these conditions the 5'-end was only partially released and the 5'-end containing dinucleotide was still present. Complete T₂ RNase digestion, however, showed only one spot because this enzyme is an endonuclease whose activity is unaffected by modified bases (see Fig. 12, digest 11b). Partial digestion of the fragment with SVP for two-dimensional homochromatography would not release the 5'-end unless the digestion was run at 45° C rather than the usual 37° C. Nuclease P₁ was incapable of releasing the 5'-end under any conditions we tried. These results would not be expected for ψUAG. Furthermore, a sequence analysis of the partial SVP digestion products of this fragment by one-dimensional high voltage electrophoresis at pH 1.9 run alongside ψψAG isolated from mammalian tRNA^phe showed
identical patterns (see Fig. 14).

A base composition analysis of spot D showed the presence of both \(^2\)m^2G and \(\psi\), therefore, it contained the needed fragment. However, in the two-dimensional TLC system we used (described above), the \(T_1\) RNase generated fragment, CUCAG, co-chromatographs with \(\text{Cm}^2\psi\text{AG}\) (78). Furthermore, the resolution of our TLC system was rather poor, and there was overlap between spots C and D. Base analysis of spot C showed it to be the \(\psi\text{CG}\) fragment. As Table 4 shows, the base composition of spot D gives what is to be expected of the three fragments \(\text{Cm}^2\text{G}\psi\text{AG}\), CUCAG, and \(\psi\text{CG}\). Fig. 7b shows the fluorograph of the base composition analysis of spot D.

III. Sequencing of a \(T_1\) RNase Fragment Utilizing Chemical Tritium Labelling

Base composition analysis of spot G (see Fig. 8 and Table 4) indicated that the \(T_1\) RNase fragment previously sequenced and designated as \(^1\text{AUCCCG}\) by \([^{32}\text{P}]-\text{labelling techniques actually contained only two unmodified C's and one m}^5\text{C. This fragment illustrates a disadvantage of \([^{32}\text{P}]-\text{labelling techniques. When a modified base occurs within an oligonucleotide, it is sometimes impossible to distinguish it from an unmodified base. In this case \([^{32}\text{P}]-\text{labelling techniques gave no clue as to which C in this sequence was modified.}

This problem was solved utilizing Randerath's tritium incorporation sequencing technique (22,32). This method has the advantage of identifying each residue unambiguously. It consists of the following steps:
Fig. 8. Fluorograph of the base composition analysis by Randerath's tritium derivatization technique (31) of spot G, m$^1$AUm$^5$ CCCG.
1) Partial digestion of the oligonucleotide fragment with snake venom phosphodiesterase and alkaline phosphatase

2) Treatment of time aliquots with sodium periodate to create oligonucleotide-3' dialdehydes

3) Reduction of the oligonucleotide-3' dialdehydes with potassium borohydride to create tritium labelled oligonucleotide-3' dialcohols

4) Resolution according to chain length of the labelled oligonucleotide derivatives by one dimensional PEI-cellulose TLC

5) Visualization of the resolved oligonucleotide derivatives by fluorography

6) In situ digestion with T2 RNase to release the labelled 3'-terminal nucleoside trialkohols

7) Elution of 3'-terminal nucleoside trialkohols

8) Identification of the labelled nucleoside trialkohols by co-chromatography with cold nucleoside trialkohol marker compounds in the two-dimensional cellulose TLC system used for base composition analysis

The experimental details of Randerath's tritium sequence analysis of oligoribonucleotides including our modified elution procedure have been described in Materials and Methods.

Figure 9-I shows the fluorograph of the PEI-TLC of labelled oligonucleotide-3' dialcohols (generated as described above) of the spot G (see Fig. 8) fragment. Fig. 9-II shows the fluorographs of the released 3'-labelled nucleoside trialkohols subjected to base composition analysis of rows a, b, c, and d. From these results the sequence $m^1$AU$m^5$CCCG could be directly deduced.
Fig. 9. (I) The fluorograph of the one-dimensional PEI-TLC of tritium labelled oligonucleotide-3' dialcohols of spot G, m\textsuperscript{1}AUm\textsuperscript{5}CCCG, generated as described in text. (II) Fluorographs of the base composition analyses of the tritium labelled nucleoside trialcohols released by \textit{in situ} digestion by T\textsubscript{2} RNase from the 3'-ends of oligonucleotides in rows a, b, c, and d of fluorograph I. The encircled spots indicate the position of non-radioactive nucleoside trialcohol uv markers. The radioactive spots in fluorographs b and c could be seen by eye, but were too light to be reproduced. In both these fluorographs a single spot was seen which co-chromatographed with the cytidine nucleoside trialcohol uv marker. These fluorographs, therefore, indicate that row a contains the undigested (by SVP) oligonucleotide since it contains guanosine at its 3'-end, and that m\textsuperscript{5}C is located at the 4th position from the 3'-end giving the oligonucleotide a sequence of m\textsuperscript{1}AUm\textsuperscript{5}CCCG.
I.

guanosine, row a

cytidine, row b

cytidine, row c

m5-cytidine, row d

II.
IV. **Sequencing of Oligonucleotides Utilizing 5'-End Group [32P]-labelling**

The sequence of *Euglena gracilis* cytoplasmic tRNA\(^{\text{phe}}\) (as well as that of the chloroplast tRNA\(^{\text{phe}}\)) was primarily deduced by [\(^{32}\)P]-post-labelling techniques.

Because of the relatively strong β particle energy of [\(^{32}\)P] [1.71 MeV (107)], these methods are highly sensitive. As little as 0.5 µg of RNA are needed for a T\(_1\) or pancreatic RNase digestion to be followed by 5'-end group [\(^{32}\)P]-labelling. Furthermore, the high resolution fractionating techniques originally developed for [\(^{32}\)P]-labelling are utilized for the labelled oligonucleotides.

A. **Base Specific Digests, 5'-End Group [32P]-Labelling with Polynucleotide Kinase, and Fingerprinting**

The [\(^{32}\)P]-post-labelling procedures we used were originally introduced by Sanger and group (90), and modified to present form by RajBhandary's group (73,75). The experimental details have been described in Materials and Methods. The following is a general sequence of the steps used:

1) Base specific digestion with either T\(_1\) or pancreatic RNase digestions of either the whole tRNA or of half-molecules (m\(^7\)G cleavage products) to produce oligonucleotide-3' phosphates

2) Removal of the 3' phosphates with bacterial alkaline phosphatase

This step was introduced to facilitate the comparison of fingerprint patterns of the post-labelled [\(^{32}\)P]-fragments with those of the \textit{in vivo} [\(^{32}\)P]-labelled fragments with which those fingerprints were first studied (23). By removing the 3' phosphates, the final
$[^{32}\text{P}]$-labelled oligonucleotides contained only a single terminal phosphate at the 5'-end. The $\textit{in vivo}$ $[^{32}\text{P}]$-labelled oligonucleotides, produced by either $T_1$ or pancreatic RNase digestions without bacterial alkaline phosphatase treatment, contained a single terminal phosphate at the 3'-end. Since there is little difference in mobility between oligonucleotides containing a single terminal phosphate at the 3' or the 5'-end, fingerprints of $[^{32}\text{P}]$-post-labelled oligonucleotides were directly comparable with those of $\textit{in vivo}$ $[^{32}\text{P}]$-labelled oligonucleotides.

3) Denaturation of the alkaline phosphatase by boiling in the presence of nitrilotriacetic acid, a chelating agent which removes zinc from the enzyme. This step is essential because active phosphatase would remove the $[^{32}\text{P}]$-label.

4) The $[^{32}\text{P}]$-labelling of the 5'-end of the oligonucleotides utilizing $\gamma[^{32}\text{P}]-\text{ATP}$ and $T_4$ induced polynucleotide kinase.

5) Removal of excess $\gamma[^{32}\text{P}]-\text{ATP}$ by conversion to $[^{32}\text{P}]-\text{glucose-6-phosphate}$ with hexokinase and D-glucose.

Glucose-6-phosphate moves more rapidly in the second dimension of the subsequent step than the oligonucleotides, whereas ATP migrates to a location within the oligonucleotide region. Therefore, this is a means of channeling away excess radiation from the area of the fingerprints containing the oligonucleotides.

6) Two-dimensional electrophoresis of the labelled oligonucleotides—"fingerprinting"

The first dimension was on cellulose acetate strips at pH 3.5. Cellulose acetate has a low capacity for material; therefore, it first
became useful with the advent of $[^{32}\text{P}]$-labelled nucleic acids (23). Since cellulose acetate has no ion exchange properties, mobility in this dimension is net charge dependent.

The second dimension was on DEAE paper at pH 1.9. Fractionation in this dimension is due to both ion-exchange and electrophoretic effects. The positive charges on the DEAE-paper cause an electro-endosmotic flow of buffer from the cathode to the anode carrying the nucleotides through the paper, and, therefore, subjecting them to ion exchange chromatography. Separations in this dimension are superior to simple chromatography on DEAE-paper; therefore, electrophoresis superimposed on chromatography is an important factor (108). Mobility in this dimension is due to both net charge and molecular weight.

The autoradiogram of this two-dimensional electrophoresis is called a fingerprint because a RNA molecule of unique sequence has a unique pattern. The mobility of oligonucleotides in this system is extremely sensitive to their base composition and sequence. Fig. 10 shows the fingerprints of A) the $T_1$ RNase digestion and B) the pancreatic RNase digestion of *Euglena gracilis* cytoplasmic tRNA$^{\text{phe}}$. The spot numbers on the fingerprints will be referred to throughout the remainder of this dissertation with the exception of tll (in most of our $T_1$ RNase digestions the oligonucleotide tll was further cleaved into two oligonucleotides, tlla and tllb).

The probable composition of an oligonucleotide may be determined by its position on the fingerprint. Brownlee (108) has
Fig. 10. Autoradiographs of fingerprints obtained by two-dimensional electrophoresis of 5'-end $^{32}$P-labelled oligonucleotides of Euglena gracilis cytoplasmic tRNA\textsubscript{Phe} obtained from (a) a T\textsubscript{1} RNase digestion and (b) a pancreatic RNase digestion. Dashed circle represents the location of the xylene cyanol dye.
null
diagrammed the relationship between the composition of an oligonucleotide and its position on the fingerprint. Certain generalizations may be made. For example, the nucleotide which has the greatest inhibitory effect on mobility in the second dimension is U, so that the fingerprint may be regarded as being composed of three sections representing oligonucleotides containing two or more, one, or no U residues. Because of the varying effects of modified nucleotides on mobility and the possibility of overlaps containing different base compositions, the usefulness of Brownlee's relationships for this project was only of a confirmatory nature.

Autoradiography of fingerprints, and the determination of molar ratios by Cherenkov counting of cutouts of the fragments were discussed in Methods. Table 5 gives the molar ratios of the $T_1$ and pancreatic RNase generated oligonucleotides of *Euglena gracilis* cytoplasmic tRNA$^{phe}$.

B. 5'-End Group Determination

An essential initial step for the sequence determination of the recovered $[^{32}P]$-labelled oligonucleotide was the identification of its 5'-end residue. This was accomplished by enzymatic degradation of the oligonucleotide to the mononucleotide level followed by co-chromatography on cellulose TLC plates with a mixture of standard uv markers as described in Methods. The 5'-terminal nucleotide, being the only one containing a $[^{32}P]$-label, was visualized by autoradiography and identified by its relative location to the standard uv markers.
Table 5. Molar Ratios$^a$ of Oligonucleotides of *Euglena gracilis* 
Cytoplasmic tRNA$^\text{Phe}$

<table>
<thead>
<tr>
<th>T$_1$ RNase Digest</th>
<th>Molar Ratio</th>
<th>Pancreatic RNase Digest</th>
<th>Spot No.</th>
<th>Molar Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>t1</td>
<td>0.63</td>
<td>p1</td>
<td>1.14</td>
<td></td>
</tr>
<tr>
<td>t2</td>
<td>0.82</td>
<td>p2</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td>t3</td>
<td>3.10</td>
<td>p3</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td>t4</td>
<td>0.76</td>
<td>p4</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>t5</td>
<td>0.77</td>
<td>p5</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>t6</td>
<td>1.00</td>
<td>p6</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>t7</td>
<td>0.10</td>
<td>p7</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>t8</td>
<td>0.10</td>
<td>p8</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>t9</td>
<td>1.20</td>
<td>p9</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>t10</td>
<td>1.00</td>
<td>p10</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>t11a and t11b</td>
<td>1.09</td>
<td>p11</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td>t12</td>
<td>0.36</td>
<td>p12</td>
<td>1.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>p13</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>p14</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>p15</td>
<td>0.50</td>
<td></td>
</tr>
</tbody>
</table>

$^a$One reason for deviations from the expected molar ratio occur when a particular oligonucleotide is not efficiently labelled by the polynucleotide kinase reaction. This is especially evident with oligonucleotides which have certain modified bases at their 5'-ends. For example, the low molar ratios on t7 and t8 are due to the presence of m$^7$G and D respectively at the 5'-ends of these two oligonucleotides.
Each oligonucleotide was digested independently with two enzymes, and each of these digests was in turn run independently in two different solvent systems, thus creating four chromatographs. Any one digest-solvent system combination might not sufficiently distinguish between certain nucleotides. By subjecting each oligonucleotide to all four systems the 5'-end could be unambiguously identified. As described in Methods, the enzymes used were T₂ RNase which releases a 5' $[^{32}P]$ nucleoside 3',5'-diphosphate ($^pXp$), and SVP or Nuclease P₁ which release a 5' $[^{32}P]$ nucleoside ($^pX$). Each digest was run with appropriate uv markers, pXp or pX. The solvents were solvent (b) and solvent (f). Fig. 11 shows the standard maps used for identifying nucleotides in the four different systems. The nucleotides represented by dashed spots are the uv standards of the four common nucleotides. Modified nucleotides are identified by their relative locations to these standards in the four systems. These maps were developed by A. Gillum and U. L. RajBhandary (75).

Fig. 12 contains the autoradiographs of the 5'-end analysis of oligonucleotides derived from the T₁ RNase digestion of cytoplasmic tRNA$^{\text{Phe}}$.

Table 6 gives the 5'-end group assignments made on the basis of these autoradiographs. It, also, gives the 5'-end groups of pancreatic RNase generated oligonucleotides obtained in a similar experiment by James Schnabel.
Fig. 11. Standard maps showing the mobilities of 5'-[^32P]-labelled modified mononucleotides relative to the mobilities of pA, pG, pC, and pU·uv markers for complete snake venom phosphodiesterase digestions (CSV) and pAp, pGp, pCp, and pUp uv markers for complete T2 RNase digestions. Dashed circles indicate the uv markers. Standard map a was run in solvent system (f). Standard map b was run in solvent system (b).
Fig. 12. Autoradiographs of the 5'-end group analyses of the complete T₁ RNase generated oligonucleotides of Euglena gracilis cytoplasmic tRNA³phe. These were obtained as follows: (a) complete SVP digestion run with solvent (f), (b) complete T₂ RNase digestion run with solvent (f), (c) complete SVP digestion run with solvent (b), (d) complete T₂ RNase digestion run with solvent (b). Circles represent locations of uv markers whose identity can be derived by referring to Fig. 11.
Table 6. 5'-End Groups of Oligonucleotides of *Euglena gracilis*
Cytoplasmic tRNA^Phe^

<table>
<thead>
<tr>
<th>T1 RNase Digest</th>
<th>5'-End Nucleotide</th>
<th>Pancreatic RNase Digest</th>
<th>5'-End Nucleotide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t1</td>
<td>C</td>
<td>p1</td>
<td>A</td>
</tr>
<tr>
<td>t2</td>
<td>C</td>
<td>p2</td>
<td>G</td>
</tr>
<tr>
<td>t3</td>
<td>A</td>
<td>p3</td>
<td>A</td>
</tr>
<tr>
<td>t4</td>
<td>ψ</td>
<td>p4</td>
<td>G</td>
</tr>
<tr>
<td>t5</td>
<td>m^1 A</td>
<td>p5</td>
<td>m^2 G</td>
</tr>
<tr>
<td>t6</td>
<td>C</td>
<td>p6</td>
<td>G</td>
</tr>
<tr>
<td>t7</td>
<td>m^7 G</td>
<td>p7</td>
<td>G</td>
</tr>
<tr>
<td>t8</td>
<td>D</td>
<td>p8</td>
<td>A</td>
</tr>
<tr>
<td>t9</td>
<td>T</td>
<td>p9</td>
<td>A</td>
</tr>
<tr>
<td>t10</td>
<td>A</td>
<td>p10</td>
<td>m^6 A</td>
</tr>
<tr>
<td>t11a</td>
<td>C</td>
<td>p11</td>
<td>A</td>
</tr>
<tr>
<td>t11b</td>
<td>ψ</td>
<td>p12</td>
<td>G</td>
</tr>
<tr>
<td>t12</td>
<td>A</td>
<td>p13</td>
<td>G^m</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p14</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p15</td>
<td>G</td>
</tr>
</tbody>
</table>
C. Sequencing by Analysis of 5' $^{32}$P-Labelled Oligonucleotides

Once the 5'-end group of the $[^{32}$P] labelled oligonucleotide was determined (and knowing that the 3'-end group was either a G for T₁ digests or a pyrimidine for pancreatic digests) the succeeding residues in the 5' to 3' direction were determined by analysis of the partial digestion products of the oligonucleotides as described in Methods.

Partial digestion of oligonucleotides was done with either snake venom phosphodiesterase (109) or Nuclease P₁ (110) as described in Methods. Reaction conditions can be adjusted so that all successive digestion products can be obtained by inactivating the enzyme during appropriate time aliquots. The electrophoretic and chromatographic techniques described below for the sequence analysis of the partial digestion products require the undigested oligonucleotide as well as all degradation products containing the 5'-end $[^{32}$P]-label. This includes the 5'-end mononucleotide even though it has been previously identified. For example, in order to sequence the theoretical oligonucleotide *pApBpCpD-OH, the oligonucleotides *pApBpCpD-OH, *pApBpC-OH, *pApB-OH, and the mononucleotide *pA-OH are needed.

Snake venom phosphodiesterase (109) is an exonuclease which recognizes the 3'-hydroxyl end of oligonucleotides and removes successively nucleotide-5'-phosphate residues in the 3' to 5' direction. An occasional difficulty arises in the use of SVP in that certain modified bases inhibit the activity of this enzyme (111)
causing build-ups of the degradation product which contains the blocking base at its 3'-end with little or no accumulation of products shorter than this. This problem could sometimes be overcome by raising the reaction temperature to 45° C instead of the usual 25° C. This method was used, for example, in the sequencing of the oligonucleotide *pψψAG. At room temperature only *pψψ-OH, *pψψA-OH, and *pψψAG-OH could be isolated. After two hours at 45° C, *pψ-OH was recovered (see Fig. 21). This technique, however, did not work for all problem oligonucleotides.

Nuclease P₁ (110) has been found to be a useful alternative nuclease for the degradation of those oligonucleotides which contain 3' or internal modified residues which hinder the action of SVP. Like SVP, Nuclease P₁ cleaves phosphodiester bonds, leaving 3'-hydroxyl and 5'-phosphate ends. Therefore, its degradation products can be analyzed identically to those of SVP. Unlike SVP, nuclease P₁ is a random endonuclease which can often cleave bonds which SVP cannot. Using normal reaction conditions (see Methods) the whole range of digestion products is usually available in a few minutes. Less than optimal cleavage of certain residues has been noted for Nuclease P₁ such as between stretches of pyrimidines. Therefore, it was often found useful to do both a SVP and nuclease P₁ digestion independently on a particular oligonucleotide, combine the degradation products of both, and proceed with sequence analysis.

For sequence analysis, the complete set of partial digestion products of a 5'-end [³²P]-labelled oligonucleotide was subjected to
either one-dimensional electrophoresis on DEAE-paper (24) or two-dimensional homochromatography (104) as described in Methods, and the mobility of its components were compared. The mobility of an individual oligonucleotide in an electrophoretic or chromatographic system is dependent on all its component nucleotides. When one component is removed (as is done successively from the 3'-end during partial digests) the remaining oligonucleotide undergoes a characteristic change in mobility. Fractionation of the entire set of partial digestion products in these systems thus allows determination of each successive nucleotide lost.

The theoretical basis for sequence analysis utilizing one-dimensional electrophoresis at pH 1.9 or 3.5 on DEAE-cellulose paper was described by Sanger (23) in 1965. This method of analysis depends on the vertical distance between successive degradation products which is actually a function of the nucleotide which has been lost. Each nucleotide has a characteristic effect expressed as its "M" value. If the distance from the origin of a parent oligonucleotide, N(n), is x and the distance from the origin of its first degradation product, N(n-1), is y, then \( M = \frac{y - x}{x} \). The M value thus calculated identifies the 3'-terminal nucleotide in the parent oligonucleotide, N(n). The 3'-terminal of N(n-1) is in turn characterized by measuring x as the distance from N(n-1) to the origin and y as the distance of N(n-2) to the origin and so on. Table 7 gives a list of the characteristic M values for the four major nucleotides at pH 3.5 and 1.9 which were published by Sanger (23).
Table 7. Published M Values for One-Dimensional Electrophoresis on DEAE-Paper at pH 3.5 or 1.9 (23,73)

<table>
<thead>
<tr>
<th>3'-Terminal Nucleotide Removed</th>
<th>M Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pH 3.5</td>
</tr>
<tr>
<td>pC</td>
<td>0.6-1.2</td>
</tr>
<tr>
<td>pA</td>
<td>2.1-2.9</td>
</tr>
<tr>
<td>pU</td>
<td>1.7-1.9</td>
</tr>
<tr>
<td>pG</td>
<td>2.6-4.4</td>
</tr>
</tbody>
</table>
This technique has certain limitations. The ranges of M values for U, A, and G at pH 1.9 and U and G at pH 1.9 are close or overlap. Also, occasional deviations from the M values have been observed (73). It is, therefore, often useful to run a sample at both pH's.

M values have been found to be meaningless for oligonucleotides whose mobilities are beyond that of the xylene cyanol tracking dye. The xylene cyanol dye usually migrates to a position between the tri- and di-nucleotides. Therefore, chromatography of the unknown alongside dinucleotides of known sequence and 5'-phosphate mononucleotide uv markers is useful.

Large oligonucleotides do not move far from the origin; crowding near the origin makes the sequencing of any oligonucleotide larger than about nucleotides impractical by this technique.

The effects of modified bases on mobility shifts in this system vary and M values have not been established for them. When an oligonucleotide contains 3'-terminal or internal modified bases it is useful to run it alongside a homologous oligonucleotide of known sequence.

Fig. 13 shows the autoradiograms of the sequence analyses by one-dimensional DEAE-paper electrophoresis at pH 3.5 of the partial SVP digests of $T_l$ fragments $t_l'$, $^*p_{CACC-OH}$ and $t_l$, $^*p_{CACCA-OH}$. [The $T_l$ fragment $t_l'$, which was isolated from the immediate lower right of $t_l$, was probably present because the purification of cyt. tRNA$^{phe}$ did not remove those tRNAs which were partially degraded by nucleotidyl
Fig. 13. Autoradiograph of the sequence analysis by one-dimensional DEAE-paper electrophoresis at pH 3.5 of the partial SVP digests of t1', *pCACC and t1, *pCACCA. Dashed circles represent location of the xylene cyanol dye. Full circles represent locations of the various nucleotide uv markers. Numbers in parenthesis refer to M values calculated as described in text.
contain.

• pCA

•  "pCA

(B

• H

• pCAC

♦ I

pCAC (0.6)

• pCACC • pCACC (0.8)

jn •

• 4  » pCACCA (2.2)

t , t

tr , »  1 6

*lo  H O  o  f | l  5 (»  W  «  «  (to  |
transferase, the enzyme which removes and replaces the terminal CCA-OH of tRNAs and thus controls their activity (121,122). Partial digestions were by SVP. M values are given in parentheses.

Fig. 14 illustrates a use of one-dimensional DEAE-paper electrophoresis which is not dependent on M values. The problems associated with the sequencing of tllb, ψAG have been discussed (see pg. 71). The presence within the oligonucleotide of the modified base, ψ, is best demonstrated by running the sample alongside the same fragment from a known source. In this experiment ψψAG was isolated from T₁ RNase fingerprint of beef liver tRNA\textsuperscript{phe} (tllb-fingerprint C, Fig. ) and analyzed alongside the same fragment from Euglena cyt. tRNA\textsuperscript{phe}. The mobilities of the partial SVP digest degradation products were identical for both beef liver and cyt. tRNA\textsuperscript{phe}s. This information plus the base composition analysis discussed previously were considered sufficient to make the sequence assignment.

Fig. 15 contains the autoradiograms of the partial SVP digests of the T₁ fragments t₂, t₃, t₆, and t₄. Of these only t₃, the dinucleotide \textsuperscript{*}pAG is here definitively sequenced. The autoradiogram shows it to be a dinucleotide with a 5'-terminal A, and since the oligonucleotide is a T₁ fragment, the sequence must be AG. Fragment t₄ is shown to be a trinucleotide with a U-like 5'-terminal. [Previous 5'-terminal analysis (see Fig. 12) had shown the 5'-terminal to be ψ]. But because the dinucleotide traveled beyond the xylene cyanol dye, calculation of a M value cannot be used to determine the
Fig. 14. Autoradiograph of the sequence analyses by one-dimensional DEAE-paper electrophoresis at pH 1.9 of the partial SVP digest of (a) tllb, *pψAG and (b) its homologous counterpart from beef liver tRNAPhe run side by side to verify that the two oligonucleotides have identical sequences. The dashed circle gives the location of the xylene cyanol tracking dye. The full circles give location of the mononucleotide uv markers.
Fig. 15. Autoradiographs of the sequence analyses by one-dimensional DEAE-paper electrophoresis at pH 3.5 of the partial SVP digests of t2,  *pCCG; t3, *pAG; t6, *pCUCAG; and t4, *pψCG. The dashed circle gives the location of the xylene cyanol tracking dye. The full circles give the locations of the mononucleotide uv markers. The numbers in parenthesis for the t6 analysis are the M values calculated as described in the text.
internal residue. Similarly, fragment t2 is seen to be a tri-nucleotide with a 5' terminal C residue, but the internal residue cannot be deduced. The fragment t6 is seen to be a pentanucleotide with a 5' terminal C, and by M value calculation CAG-OH on the 3'-end. Once again the 3'-terminal of the dinucleotide degradation product cannot be directly deduced. One way to determine the identity of the 3'-terminals of these dinucleotides would be to run them alongside known standards. Instead we used two-dimensional homochromatography to solve these problems, and as our major method of sequencing.

Two-dimensional homochromatography was first applied to DNA sequencing (95,114,115) and more recently has been adapted to RNA sequencing (73,116). It has many advantages over one-dimensional electrophoresis, and was the primary method of sequencing used in this project. Instead of depending on the often ambiguous M values, sequences can be directly read by the characteristic mobility shifts which homologous series of degradative products display in this two-dimensional system. Furthermore, large oligonucleotides (as large as 15 or more) can be sequenced using this method. Experimental details have been given in Methods.

The first dimension is on cellulose acetate strips at pH 3.5. The mobility of an oligonucleotide in this dimension is dependent on the unique charge and molecular weight contribution of its components. The relative mobility of a series of homologous partial degradation products is, therefore, a function of the nucleotide by which they differ. The removal of a 5'-nucleotide results in increased or
decreased electrophoretic mobility for the remaining oligonucleotide, and this difference in mobility is a function of the average charge of the nucleotide removed. The average charge of each nucleotide can be calculated knowing its pKa using the Henderson-Hasselbach equation, \( \text{pH} = \text{pKa} + \log \left( \frac{\text{base}}{\text{acid}} \right) \), and assuming the contribution of the phosphate is -1 at this pH. Table 8 lists the pKa's of the four major 5'-nucleotides and their calculated average charge at pH 3.5. The net mobility contribution of a nucleotide to its parent oligonucleotide is a result of charge and mass effects where negative charge increases mobility and mass had a drag effect (73). The presence of the nucleotide pU, which has the largest negative charge, causes the greatest increase in the electrophoretic mobility of its parent oligonucleotide. (The removal of pU, therefore, causes the greatest decrease in mobility.) The effects of the presence and removal of pG resemble that of pU except that they are less extreme because pG has a slightly smaller net negative charge. On the other hand the presence of pC with its very small negative charge usually slows its parent oligonucleotide's mobility because its mass contributes more drag than its charge can compensate for. An exception occurs when the parent oligonucleotide is only slightly charged, and the small charge of pC makes a substantial increase in total charge. For example, the mobility of pCC is slightly greater than that of pC (73). The effect of pA on the mobility of its parent oligonucleotide is minimal because its partial net negative charge effectively balances the drag effect due to its mass.
The second dimension of this system is homochromatography (76) on DEAE-cellulose thin layer plates at pH 4.7. Homochromatography is a type of displacement chromatography where the RNA fragments in the homomix (see Methods) displace similarly sized fragments on the plate resulting in a fractionation based on size with smaller fragments traveling further than larger fragments. The removal of a terminal purine residue causes a larger mobility shift than the removal of a terminal pyrimidine due to differences in mass. This differential vertical distance depending on class of nucleotide removed is illustrated in Fig. 16. This phenomenon provides a convenient confirmation of sequence assignments based on the horizontal component of angular mobility shifts. For example, if a homochromatograph shows a typical U "jump" in the horizontal direction between an oligonucleotide and its first degradation product in a homologous series, the vertical component should have a mobility shift which has a pyrimidine character.

It should be emphasized that the removal of a nucleotide from its parent oligonucleotide has the opposite effect on electrophoretic mobility as its presence had. The effects of the removal of the four major 5' nucleotides on the typical oligonucleotide in terms of angular mobility shifts in the two dimensional homochromatography system are summarized in Fig. 16.

It should be noted that modified nucleotides do not necessarily exhibit the same characteristic angular mobility shifts as do their parents upon removal. Those modifications which result
Table 8. The pKa's (47) and Average Charge (pH 3.5) of the 5'-Monophosphate Nucleotides

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>pKa</th>
<th>Average Charge (7), pH 3.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>pU</td>
<td>-</td>
<td>-1</td>
</tr>
<tr>
<td>pG</td>
<td>2.4</td>
<td>-0.93</td>
</tr>
<tr>
<td>pA</td>
<td>3.74</td>
<td>-0.34</td>
</tr>
<tr>
<td>pC</td>
<td>4.5</td>
<td>-0.10</td>
</tr>
</tbody>
</table>

Fig. 16. Effects of removal of nucleotides on electrophoretic mobility of parent oligonucleotides in two-dimensional homochromatography (angular mobility shifts). Vertical differences (second dimension) in mobility between removal of purines (pu) and pyrimidines (py) (73,116).
in the change of the net charge of the nucleotide especially show unique angular mobility shifts. As more and more work is being done using this technique, a library of characteristic shifts for those nucleotides which have unique shifts is being built up. On the other hand, certain modified bases show little or no detectable difference of angular mobility shifts from their parent nucleotides. For example, a U "jump" cannot be distinguished from a D, T, or ψ "jump."

This, in fact, points out the major weakness in the two-dimensional homochromatography technique, i.e. the technique does not provide unambiguous identification of each residue and often corroboratory evidence from base composition analysis, one-dimensional electrophoresis alongside known standards, or even a different sequencing technique such as Randerath's chemical tritium labelling method (22) is needed to make a final sequence assignment.

A special problem arises when an oligonucleotide contains a residue which is sensitive to pH changes encountered in processing. For example, under alkaline conditions such as are necessary for snake venom phosphodiesterase digestions, m^1A partially isomerizes to m^6A (117,118). This leads to two sets of digestion patterns thus making the reading of angular mobility shifts a matter of guess work. Methods have been developed for dealing with such problem oligonucleotides and are discussed in the following section.

Deviations from characteristic angular mobility shifts occur in long fragments which, because of their great mass, are less responsive to the removal of a single nucleotide. This results in
restricted jumps which are more difficult to interpret. This puts an upper size limit of 20 or so residues on the length which can be effectively sequenced.

Before actual analysis of the oligonucleotide by two-dimensional homochromatography, the kinetics of the digestion were usually tested by one-dimensional homochromatography of samples of each time aliquot as described in Methods. This procedure facilitated the pooling of the appropriate time aliquots as well as quantities needed so that the two-dimensional homochromatography would display each digestion intermediate in proportions as equal as possible.

Fig. 17 is the autoradiograph of a typical one-dimensional homochromatography check, in this case, of the oligonucleotide \( ^* \text{pCUCAG-OH} \). At zero time (before addition of SVP) only a single spot, the parent oligonucleotide, is seen. After enzyme addition, the various time aliquots display varying amounts of each digestion intermediate. It can be seen that by pooling the zero, 20, and 160 minute time aliquots, a representative display of all digestion intermediates can be fractionated in the two-dimensional system.

Most of the \( T_1 \) and pancreatic RNase fragments of \textit{Euglena gracilis} cytoplasmic tRNA\(^{\text{phe}}\) (as well as those of the chloroplast tRNA\(^{\text{phe}}\)) were sequenced using two-dimensional homochromatography. The rest of this section will be devoted to displaying the two-dimensional homochromatography of these \( T_1 \) fragments which were readily sequenced using this technique. [Jim Schnabel (27) sequenced most of the pancreatic RNase fragments, and his results are
Fig. 17. Autoradiograph of the one-dimensional homochromatography of the partial SVP digestion of \*pCUCAG showing the extent of the digestion at various times after enzyme digestion. (Zero time is before addition of SVP to the oligonucleotide.) Circles show location of the xylene cyanol dye.
given in Table 9]. The following section will be devoted to "problem" oligonucleotides, and their sequencing.

The two-dimensional homochromatographs shown below were obtained using the standard conditions described in Methods. Since the xylene cyanol tracking dye travels faster than pC and pA and slower than pG and pU in the first dimension, the positioning of the DEAE plate on the cellulose acetate strip depended on the 5'-terminal (previously determined) of the oligonucleotide. If the oligonucleotide had a pG or pU 5'-terminal the xylene cyanol dye spot on the cellulose acetate strip was positioned close to the left hand edge of the DEAE plate, since it could be assumed that all the other components of the digestive series moved faster than the dye. If the 5'-terminal was pC or pA, the xylene cyanol dye was centered on the plate, since some components would move faster than the dye and some slower. This explains the position of the xylene cyanol spots on the homochromatographs shown which are symbolized by a B (for blue) encircled by dashed lines. Each homochromatograph will be displayed alongside an explanatory diagram which traces the angular mobility shifts and gives their interpretation as to which nucleotide is represented by this shift. If one reads from the spot which represents the 5'-terminal (labelled pX where X is a nucleotide) toward the undigested parent oligonucleotide (the lowest right hand spot), each shift can be considered as the addition of the displayed nucleotide. Reading in the opposite direction, each shift can be considered to be the removal of the displayed nucleotide.
Fig. 18 shows the two-dimensional homochromatograph of a partial SVP digest of t2, *pCCG which was displayed earlier in this section (see Fig. 15) as being analyzed for sequence by one-dimensional electrophoresis, but, as explained, not unambiguously. Here the assignment is confirmed. As explained above the mobility of pCC is actually slightly greater in the first dimension than the mobility of pC because the total negative charge, though slight, is relatively greatly increased.

Fig. 19 shows the two-dimensional homochromatograph of the partial SVP digest of another trinucleotide, t4, *pψCG which, also, was displayed earlier in a one-dimensional electrophoretic analysis which was not conclusive. Once again the sequence is here confirmed in this system. It is interesting to compare Fig. 18 and Fig. 19. In the latter the addition of a pC slows the oligonucleotide's mobility in the first dimension rather than increasing it like in the former. This is because the additional charge added by pC is so slight compared to the full charge of pψ that the additional mass added by pC is the overriding factor rather than charge.

Fig. 20 and Fig. 21 show the two-dimensional homochromatographs of partial SVP digests of fragments t1l and t1lb. As explained earlier (see p. 71) these fragments *pCm2GψψG and *pψψG respectively are the result of different T1 RNase digestions which due to some difference in reaction conditions either were able to or could not cleave m2G. It may be noted that the angular mobility shifts of the spots to the right of the blue dye in Fig. 20 are similar to those in
Fig. 18. Autoradiograph and explanatory tracing of the two-dimensional homochromatography of the partial SVP digestion products of t2, *pCCG. Dashed circle shows location of the xylene cyanol dye.
Fig. 19. Autoradiograph and explanatory tracing of the two-dimensional homochromatography of the partial SVP digestion products of t4, *pψCG. Dashed circle shows locations of the xylene cyanol dye.
Fig. 20. Autoradiograph and explanatory tracing of the two-dimensional homochromatography of the partial SVP digestion products of tll, *pCm2GψψAG. The dashed circle shows the location of the xylene cyanol dye.
Fig. 21. Autoradiograph and explanatory tracing of the two-dimensional homochromatography of the partial SVP digestion products of tllb, *pψψAG. Dashed circle shows the location of the xylene cyanol dye.
Fig. 21 though the jumps are more restricted. This illustrates the principle described above that the longer an oligonucleotide, the less responsive it is to the loss of a single nucleotide.

Figs. 22, 23, and 24 show the homochromatographs of the partial SVP digests of fragments t6, t9, and t10 respectively. The sequence analysis of t6 by one-dimensional electrophoresis of the partial SVP digest has been previously displayed (see Fig. 15). Once again two-dimensional homochromatography was necessary for confirmation.

D. Sequencing of Oligonucleotides Containing "Problem" Modified Bases

1. Sequencing of the T1 and Pancreatic RNase Generated Anticodon-Containing Fragments Which Contain the "Y" Base

The T1 RNase generated anticodon-containing fragment of \textit{Euglena gracilis} cytoplasmic tRNA\textsuperscript{phe}, t12, occupies the same position on the fingerprint as the homologous beef liver tRNA\textsuperscript{phe} oligonucleotide. It was therefore, initially assumed that these two fragments could have similar or identical sequences. Base composition analysis of the \textit{Euglena} anticodon-containing T1 fragment confirmed that this was likely (see Table 3). When partial SVP digests of the \textit{Euglena} and beef liver fragments were run side by side in the one-dimensional electrophoresis system at pH 1.9, similar patterns were seen. However, because these fragments are 15 nucleotides long, crowding near the origin made actual sequence analysis impossible by this technique.
Fig. 22. Autoradiograph and explanatory tracing of the two-dimensional homochromatography of the partial SVP digestion products of t6, *pCUCAG. The dashed circle represents location of the xylene cyanol dye.
Fig. 23. Autoradiograph and explanatory tracing of the two-dimensional homochromatography of the partial SVP digestion products of t9, \( ^*_{p}T\Psi CG \). The dashed circle represents the location of the xylene cyanol dye.
Fig. 24. Autoradiograph and explanatory tracing of the two-dimensional homochromatography of the partial SVP digestion products of t10, *pACUUAm^2G. The dashed circle represents the location of the xylene cyanol dye.
Several problems arose associated with the sequencing of this oligonucleotide due to the presence of the "Y" base (see Fig. 26). Fingerprints which had been obtained using DEAE-paper run at pH 1.9 in the second dimension showed fuzzy, indistinct t12 spots. This was because the Y base is acid labile. Zachau (119) demonstrated that mild treatment of \( \text{tRNA}^{\text{phe}} \) with acid (pH 2.9, 37°C, 3-4 hours) will cause some excision of the Y base from the tRNA without breaking the chain. It could, therefore, be assumed that a great amount of excision occurs during a standard fingerprinting. To alleviate this problem we used the cellogel-PEI TLC fingerprinting technique (see Methods) to purify t12. This had two advantages. First, the cellogel gave more distinct, compact spots, in effect, separating the oligonucleotides still containing Y base from those within which it was degraded. Secondly, the PEI plates were run at a higher pH (pyridinium formate, pH 3.4) resulting in less degradation (see Fig. 25).

Several different structural forms of the Y base have been characterized from the \( \text{tRNA}^{\text{phe}} \)s of several species (120). The so-called "peroxy Y" base of mammalian liver and wheat germ is illustrated in Fig. 26. The Y base of Euglena cyt. \( \text{tRNA}^{\text{phe}} \) has not been characterized. If it turns out to be the "peroxy" Y of mammalian liver and wheat germ \( \text{tRNA}^{\text{phe}} \) with which it shares so many other structural features, then various chemical degradations of the base's extensive side chain can be envisioned during processing for sequencing. For example, hydrolysis of the ester groups might occur
Fig. 25. Autoradiograph of PEI-cellulose fingerprint of T₁ RNase generated oligonucleotides from *Euglena gracilis* cytoplasmic tRNA^Phe_. Dashed circle represents location of the xylene cyanol dye. This method of fingerprinting was utilized to isolate t₁2 (see text). The spot number assignments were made in the most part by comparison with the fingerprint in Fig. 10.

Fig. 26. Structure of the "peroxy" Y base of beef liver and wheat germ tRNA^Phe_ (129).
either during the second dimension of fingerprinting at pH 1.9 or during snake venom digestions at pH 8.9. Furthermore, the presence of dithiothreitol during the polynuclotidase kinase reaction would reduce the peroxyl group. It should be re-emphasized that such side reactions are conjectural since the actual structure of the Euglena cyt. tRNA\(^{\text{phe}}\) Y base has not been characterized. However, during sequence analysis by two-dimensional homochromatography the region of the homochromatograph containing the Y base was often fuzzy, or sometimes the Y base was entirely missing, thus suggesting extensive degradation.

Another problem arose due to the fact that the Y base inhibits the action of snake venom phosphodiesterase (ill). To obtain the complete series of homologous degradation products from t12 it was necessary to subject this fragment both to SVP and nuclease P\(_1\) partial digestions independently and combine the products for two-dimensional homochromatography.

A third problem associated with the sequence determination of this fragment was its length. Two-dimensional homochromatography on short plates (see Methods) gave undecipherable information in the lower region of the homochromatograph containing the longer oligonucleotides because they were too crowded. On the other hand, one such plate gave us accurate sequence information for the first ten nucleotides (starting from the 5'-end) including a typical Y jump (see Fig. 27). In order to zero in on the longer fragments, we used a long plate (see Methods) and got accurate sequence information on
Fig. 27. Autoradiograph and explanatory tracing of the two-dimensional homochromatography of the combined partial SVF and Nuclease P1 digestions of t12, *pACmUGmAYmCUAAAG. The dashed circle represents the location of the xylene cyanol dye. The second dimension was run on a short DEAE plate (20 x 20 cm).
the last eight nucleotides (terminating in the 3'-end). In this particular homochromatograph the Y base was completely missing (see Fig. 28). However, by combining the information from these two experiments a complete sequence assignment was made. With this information plus the base composition analysis (see Table 3), we concluded that the anticodon-containing T1 fragment from cyt. tRNA^phe was identical to the homologous fragment in mammalian liver tRNA^phe.

(In an additional experiment we determined that the extra spot seen in both Fig. 27 and Fig. 28 below the parent oligonucleotide is not part of the sequence by running two-dimensional homochromatography on undigested t12 fragment, and noting that this extra spot was present. It could be a breakdown product of t12, such as t12 with excised Y base.)

Similarly, the pancreatic RNase anticodon containing fragment pl3, pG^AAYAψ had caused difficulties in sequencing due to the presence of the Y base. Standard fingerprinting gave pl3 spots which were extremely diffuse. In order to get a more compact spot on the fingerprint, a cellogel strip (see Methods) was used in the first dimension rather than cellulose acetate. The spot was eluted from the fingerprint, its purity checked by one-dimensional homochromatography of an aliquot, and two partial digests—one with SVP and one with Nuclease P1—run independently. The appropriate aliquots were combined (after checking by one-dimensional homochromatography) and the pooled aliquots were subjected to two-dimensional
Fig. 28. Autoradiograph and explanatory tracing of the two-dimensional homochromatography of the combined partial SVP and Nuclease P1 digestions of t12, *pACmUGmAAYAψCUAAAG. The dashed circle represents the location of the xylene cyanol dye. The second dimension was run on a long DEAE plate (20 x 40 cm). The information obtained from this autoradiograph and the one illustrated in Fig. 27 gave us enough information to establish the entire sequence of t12.
homochromatography (with a cellogel strip in the first dimension). The resulting homochromatograph gave the expected results (see Fig. 29).

2. Sequencing of Fragments Containing $m^1A$ or $m^7G$

The presence of the modified nucleotides $m^1A$ or $m^7G$ in an oligonucleotide complicates sequencing for two reasons. First, both these nucleotides have pKa values higher than 7.0 (121), and at pH 3.5 (the pH of the first dimension in two-dimensional homochromatography) both these nucleotides have full positive charges. The presence of these nucleotides within an oligonucleotide, thus causes uncharacteristic angular mobility shifts (as will be discussed below). Therefore, previous knowledge of the existence of these nucleotides in an oligonucleotide is necessary for proper interpretation of the homochromatographs. Secondly, both $m^1A$ (67) and $m^7G$ (21) partially isomerize under alkaline conditions (such as during SVP digestions, pH 8.9) to compounds which do not contain positively charged ring systems. When isomerization has been extensive, two sets of digestion patterns appear on the homochromatograph which are indistinguishable.

Interpretation of the unusual angular mobility shifts caused by the presence of either of these nucleotides in an oligonucleotide is based on the knowledge that because of their high pKa values their presence has an opposite effect in both dimensions compared to other nucleotides. In the first dimension on cellulose acetate strips at
Fig. 29. Autoradiograph and explanatory tracing of the two-dimensional homochromatography of the combined partial SVP and Nuclease P1 products of pl3, *pGmAAAYψ. The dashed circle represents the location of the xylene cyanol dye.
pH 3.5 these two nucleotides have positive charges, and their addition to an oligonucleotide dramatically slows its mobility. The second dimension, homochromatography on DEAE-cellulose plates at pH 4.7 usually separates oligonucleotides on the basis of chain length with the shorter fragments of a series moving faster than the longer fragments. The addition of either of these nucleotides to an oligonucleotide, however, decreases its total negative charge and, therefore, makes it move faster in the second dimension. The overall result in two dimensions is that the addition of either of these nucleotides to an oligonucleotide causes it to move dramatically slower in the first dimension and faster in the second (75). This is illustrated in Fig. 30a for the p8' fragment, *pAAAGm7CD from mammalian liver tRNA^phe.

The presence of m7G or m1A at the 5'-end of an oligonucleotide also has a perceptable effect on angular mobility. The 5'-end *pm7G or *pm1A actually migrates toward the cathode (75) and, therefore, is usually not seen on the homochromatography. The entire digestion series migrates slower because of the presence of either of these nucleotides at the 5'-end, and its presence may cause uncharacteristic mobility shifts in the addition of ordinary nucleotides. This is dramatically illustrated in Fig. 31 for the t7 fragment, *pm7GUCCUG. As discussed in a previous section, the addition of C usually slows an oligonucleotide in the first dimension because the drag caused by the additional mass is not sufficiently compensated for by the slight negative charge of C. In this case
Fig. 30. Autoradiographs and explanatory tracings of the two-dimensional homochromatography of the partial SVP digestion products of the pancreatic digestion generated oligonucleotide, *pAAAGm7GD from beef liver tRNA_Phe (a) before and (b) after incubation in 50 mM ammonium bicarbonate, pH 9.0 at 85°C for 30 minutes. The dashed circles represent the location of the xylene cyanol dye.
Fig. 31. Autoradiographs of the two-dimensional homochromatography of the partial SVP digestion products of t7, *pm\(^7\)GUCCUG (a) before and (b) after incubation in 50 mM ammonium bicarbonate, pH 9.0 at 85° C for 30 minutes. The dashed circles represent the location of the xylene cyanol dye.
however, the net negative charge of $^*\text{pm}^7\text{GU}$ is so slight that the additional charge added by the addition of C's increases the mobility of the oligonucleotide. This same effect is seen in the homochromatogram of t5, $^*\text{pm}^1\text{AUm}^5\text{CCC}$, Fig.

The second problem, that of partial isomerization of $^1\text{A}$ and $^7\text{G}$ under alkaline conditions to other non-positively charged compounds was solved using a technique described by Gillum for $^1\text{A}$ (75). Noting that the isomer involved, $^6\text{A}$, was stable once produced, and furthermore, behaved like an A because it had no positive charge, Gillum incubated $^1\text{A}$ containing fragments in alkaline conditions, resolved the $^6\text{A}$ derivative from the remaining $^1\text{A}$ containing material by two-dimensional cellogel-PEI fingerprinting, and proceeded to analyze partial digests by two-dimensional homochromatography.

Fig. 32 illustrates the dramatic improvement in interpretation afforded by this technique. Here t5, $^1\text{AUm}^5\text{CCC}$, was incubated in 50 mM ammonium bicarbonate, pH 9.0 at $85^\circ\text{C}$ for 30 minutes. The $^6\text{AUm}^5\text{CCC}$ isomer thus created was isolated as the faster moving of two spots in the first dimension of a two-dimensional cellogel-PEI fingerprint. The two spots, the $^1\text{A}$ containing fragment and the $^6\text{A}$ containing fragment, were extracted (see Methods) and analyzed by two-dimensional homochromatography of combined partial SVP and Nuclease P1 digests. As we see in Fig. 32a, the 5'-terminal nucleotide is not present, the di and trinucleotides are not resolved, and the C jumps are uncharacteristic (as explained above). In
Fig. 32. Autoradiographs and explanatory tracings of the two-dimensional homochromatography of the partial SVP digestion products of t5, mAUm5CCCG (a) before and (b) after incubation in 50 mM ammonium bicarbonate, pH 9.0 at 85° C for 30 minutes. The dashed circles represent the location of the xylene cyanol dye.
homochromatograph b, the 5'-terminal $m^6A$ is clearly discernible, and all nucleotides give characteristic angular mobility jumps.

Because the imidazole ring of $m^7G$ was known to be labile under alkaline conditions creating an open ring with the bond between $C^8$ and $N^9$ broken, leaving an aldehyde and an amine function and thus removing the positive charge, we tried the same conditions described above for t5 for t7, $m^7GUCCUG$. Once again a fast and slow moving spot were isolated from the two-dimensional cellogel-PEI fingerprint. Fig. 31 shows homochromatographs of partial SVP digests of a, the slow moving $m^7G$ containing fragment, and b, the faster moving X containing fragment, where X is the open ring form of $m^7G$. Once again as in Fig. 32, we see the great improvement interpretability of the angular mobility shifts once the positive charge has been removed. As a result of this particular experiment we developed a technique which substantially increased the sensitivity of the homochromatographic technique, that is we found that a fragment with as little as 5,000 Cherenkov cpm could be sequenced. The fast moving X containing fragment contained only 9,000 cpm after extraction from the PEI plate. We dispensed with the one-dimensional homochromatographic check of the partial SVP digest. After two-dimensional homochromatography, the plate was coated with PPO by pouring a 7% solution in diethyl ether over it. Autoradiography for three days at $-80^\circ$ C produced homochromatograph b in Fig. 31.

Fig. 30 shows the homochromatographs of the mammalian liver p8' fragment, *pAAAGm7GD (a) before and (b) after alkaline incubation
and cellogel-PEI purification as described above. Homochromatograph a shows the reverse jump caused by the addition of a m^7G. As can be seen, it would be impossible to correctly interpret this homochromatograph without prior knowledge of the presence of m^7G. Once again the alkaline incubated open ring form gives an easily interpreted homochromatograph (b).

Table 9 contains a summary of the oligonucleotides in *Euglena gracilis* cytoplasmic tRNA^phe^ obtained by complete T₁ and pancreatic digestions, and sequenced utilizing the methods in this and the preceding section.

E. Creation of Overlap Sequences

1. Cleavage of the 3'-Phosphodiester Bond Adjacent to m^7G

The phosphodiester bond at the 3'-end of m^7G was chemically cleaved (19) as described in Methods, and the two half molecules thus obtained were separated by one-dimensional polyacrylamide gel electrophoresis and eluted, also as described. The 5'-half was apparently partially decomposed, and 5'-half oligonucleotide fragments contaminated the 3'-half band. Nevertheless, quantitative analysis of T₁ and pancreatic RNase fingerprings of the 3'-half band indicated which fragments were actually from the 3'-half (30 nucleotides long). These results are displayed in Fig. 36.

2. Partial Pancreatic RNase Digestion

About 1.0 A₂₆₀ unit of *Euglena* cytoplasmic tRNA^phe^ was partially digested with pancreatic RNase (18) as described in Methods.
Table 9.

<table>
<thead>
<tr>
<th>Oligonucleotides in Cytoplasmic tRNA$^{ph}$ of <em>Euglena gracilis</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T₁ RNase Digest</strong></td>
</tr>
<tr>
<td>Spot No.</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>t₁</td>
</tr>
<tr>
<td>t₂</td>
</tr>
<tr>
<td>t₃</td>
</tr>
<tr>
<td>t₄</td>
</tr>
<tr>
<td>t₅</td>
</tr>
<tr>
<td>t₆</td>
</tr>
<tr>
<td>t₇</td>
</tr>
<tr>
<td>t₈</td>
</tr>
<tr>
<td>t₉</td>
</tr>
<tr>
<td>t₁₀</td>
</tr>
<tr>
<td>t₁₁</td>
</tr>
<tr>
<td>t₁²</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*The sequence of p₁³ is G$^m$AAYAψ*
Fig. 33. Autoradiograph and explanatory tracing of the two-dimensional homochromatography of combined partial SVP and Nuclease P1 digestion products of the partial RNase fragment, *pGGAGAG[Gψ]. The dashed circle represents the location of the xylene cyanol dye.
Fig. 34. Autoradiograph and explanatory tracing of the two-dimensional homochromatography of the combined partial SVP and Nuclease P1 digestion products of the partial pancreatic RNase fragment, *pGCCGAGUU. The dashed circle represents the location of the xylene cyanol dye.
Fig. 35. Autoradiograph and explanatory tracing of the two-dimensional homochromatography of the combined partial SVP and Nuclease P1 digestion products of the partial pancreatic RNase fragment, *pAm2GCUCAGDD. The 5'-end nucleotide, *pA, is not seen in this autoradiograph. The dashed circle represents the location of the xylene cyanol dye.
The oligonucleotides were labelled at their 5'-ends with $^{32}\text{P}$ and resolved by two-dimensional gel electrophoresis (89) also as described in Methods. Analysis was by typical two-dimensional homochromatography of combined partial SVP and Nuclease $P_1$ digests. Figs. 33, 34, and 35 show the homochromatographs of three such partial pancreatic fragments. These were used as overlap fragments to help order sequences of smaller fragments obtained by complete $T_1$ and pancreatic RNase digestions. Fig. 36 shows the partial pancreatic fragments and how they helped establish the total sequence of *Euglena* cytoplasmic tRNA$^{\text{phe}}$.

3. Labelling of the 3'-End of the tRNA with $^{32}\text{P}$

Dr. S. H. Chang labelled the 3'-end of intact *Euglena* cytoplasmic tRNA$^{\text{phe}}$ by first treating it with SVP under mild conditions to remove part of its 3'-terminal CCA-OH end, and then labelling with $^{32}\text{P}$ at the 3'-end using tRNA nucleotidyl transferase in the presence of $\alpha$-$^{32}\text{P}$-ATP and non-radioactive CTP as described by Silberklang *et al.* (122). Two-dimensional homochromatography of a partial Nuclease $P_1$ digestion of this 3'-$^{32}\text{P}$-labelled tRNA established the sequence of the first 13 nucleotides from the 3'-end on. These results are displayed in Fig. 36.

F. The Nucleotide Sequence of *Euglena gracilis* Cytoplasmic tRNA$^{\text{phe}}$

The information obtained by the sequencing of complete $T_1$ and pancreatic RNase oligonucleotides (see Table 9) plus the sequencing of longer oligonucleotides created as detailed above was sufficient to
Fig. 36. Results of the sequence analysis of various oligonucleotides. The sequence at the top is that of *Euglena gracilis* cytoplasmic tRNA^phe^ in linear form. The fragments shown in (a) through (d) are large oligonucleotides produced by partial pancreatic RNase digestion. The sequence in (e) was obtained by a partial Nuclease P₁ digestion on 3'-[^32P]-labelled tRNA. The sequence in (f) was placed at the 3'-end of the tRNA by the results of the m⁷G cleavage experiment. Oligonucleotides from complete T₁ RNase digestion (g) through (j), and complete pancreatic RNase digestion (k) provided overlaps for some of the larger oligonucleotides.
pGCCGACUUAGCUCAGDDGGGAGAGGCGYGAGAC*UG AAYA*CUAAAGGUGCCUUGT*YG AUCCCGGGA*YG CGGCACCAOH

GCCGACUU
(a)

GGGAGAGGCGY
(c)

AAAGGUCCCCUGT*YG
(d)

---GGAGYC CGGCACCA
(e)

AGCUCAGDD
(b)

ACUGAAYA*CUAAAG
(j)

ACUUAAG
(g)

DDG
(h)

*YGAG
(i)

AGAC*U
(k)

---T*YG AUCCCGGGA*YG CGGCACCA
(f)
establish the entire nucleotide sequence of *Euglena gracilis* cytoplasmic \( \text{tRNA}^{\text{phe}} \). Fig. 36 shows the results of these analyses. The sequence at the top of the page is that of the total tRNA in linear form. Fragments (a) through (d) are long oligonucleotides created by partial pancreatic RNase digestion. The sequence shown in (e) was obtained by a partial Nuclease \( P_1 \) digestion on the 3'-[\( ^{32} \text{P} \)]-labelled tRNA. The sequence shown in (f) was established by analysis of the 3'-half of the tRNA resulting from \( \text{m}^7 \text{G} \) cleavage. Oligonucleotides from complete \( \text{T}_{1} \) RNase digestion (g) and complete pancreatic RNase digestion (k) are selectively shown to provide overlaps between the longer fragments.

Fig. 37 shows the *Euglena gracilis* cytoplasmic \( \text{tRNA}^{\text{phe}} \) sequence in cloverleaf form. Like all other \( \text{tRNA}^{\text{phe}} \) sequenced to date, this \( \text{tRNA}^{\text{phe}} \) contains 76 nucleotides. There are no unusual secondary structural features of this tRNA. The dihydrouridine arm contains 16 nucleotides and the variable loop contains 5 nucleotides. As shown in Fig. 37, this tRNA contains a large amount (18 total) of modified nucleosides, which is typical of eukaryotic tRNAs. Unlike mammalian \( \text{tRNA}^{\text{phe}} \), which contains a \( \text{m}^1 \text{A} \) at position 14 in the D loop, *Euglena gracilis* cytoplasmic \( \text{tRNA}^{\text{phe}} \) does not contain \( \text{m}^1 \text{A} \) at this position, but instead has a mixture of A and \( \text{m}^6 \text{A} \). Like mammalian \( \text{tRNA}^{\text{phe}} \), it contains a \( \text{m}^1 \text{A} \) at position 58 in the T\( \text{YC} \) loop. Uniquely, it contains a \( \text{m}^5 \text{C} \) at position 60 in the D loop, and a \( \psi \) at position 68 in the acceptor stem. Its anticodon is \( \text{G}^{\text{mAA}} \).
Fig. 37. The nucleotide sequence of *Euglena gracilis* tRNA^{phe} represented in cloverleaf form.
DISCUSSION

We have elucidated the complete nucleotide sequence of *Euglena gracilis* cytoplasmic tRNA\(^{\text{phe}}\) (28). Our primary reason for undertaking this project was to obtain a comparison of the sequence of this tRNA with the sequence of *Euglena gracilis* chloroplastic tRNA\(^{\text{phe}}\) which we had previously sequenced (27). This comparison was of special interest because the chloroplastic sequence was the first elucidated of an organelle tRNA, and any conclusions drawn concerning the evolutionary implications of this sequence were considered speculative without the sequence of its cytoplasmic counterpart.

Fig. 38 shows in cloverleaf form the sequence of the tRNAs\(^{\text{phe}}\) of *Euglena gracilis* from (a) the cytoplasm and (b) the chloroplast. [In the cytoplasmic tRNA\(^{\text{phe}}\) sequence U\(^*\) represents a derivative of uridine (discussed below) and A\(^*\) is a mixture of A and m\(^6\)A. Nucleosides shown in parenthesis are those present in beef liver cytoplasmic tRNA\(^{\text{phe}}\). In the chloroplastic tRNA\(^{\text{phe}}\) structure U\(^*\) is either a U or D, A\(^*\) is most probably ms\(^2\)i\(^6\)A (2-methylthio-\(N^6\)-isopentenyladenosine) and Py\(^*\) is (4abu)\(^3\)U(N\(^3\)-4(2-amino) butyryl uridine) (27)].

Unlike many tRNAs which exist in families of distinct iso-accepting species, tRNA\(^{\text{phe}}\) usually exists as a single species (34) within the system it occupies. Because all tRNAs\(^{\text{phe}}\) are in this sense homologous, evolutionary conclusions drawn from sequence comparisons of this tRNA are valid.
Fig. 38. Nucleotide sequences of *Euglena gracilis* tRNAs\textsubscript{Phe} from (A) the cytoplasm and (B) the chloroplasts shown in cloverleaf form. Nucleosides shown in parenthesis in (A) are those present in beef liver cytoplasmic tRN\textsubscript{Phe}. U* and A* in (A) and U*, A*, and Py* in (B) are explained in the text.
A comparison of the twelve tRNAs\textsuperscript{phe} sequenced to date (34) shows substantial conservation. However, the tRNAs\textsuperscript{phe} can be subdivided into two groups, prokaryotic and eukaryotic, on the basis of the presence or absence of certain modified bases and on the location of specific nucleotides within the sequence. Thus "prokaryotic" tRNAs\textsuperscript{phe} usually lack the modified bases $m^2G$, $m^2G$, Cm, Gm, $m^5C$, Y, and $m^1A$ which are considered characteristic of the tRNAs\textsuperscript{phe} of the eukaryotic cytoplasm (see Fig. 4). Some positions within the tRNA\textsuperscript{phe} have specific parent nucleotides (disregarding post-transcriptional modifications) which are characteristic of either eukaryotic cytoplasmic or prokaryotic tRNAs\textsuperscript{phe}. These are summarized in Table 10.

The \textit{Euglena} cytoplasmic and chloroplastic tRNAs\textsuperscript{phe} have a 64% sequence homology, a figure which indicates a relatively high degree of divergence. (It should be noted that calculations of sequence homology take into account only differences in the parent nucleotide occurring at homologous positions, and not differences due to post-transcriptional modifications.) Both these tRNAs like all other tRNAs\textsuperscript{phe} sequenced to date (34) contain 76 nucleotides. The chloroplastic tRNA\textsuperscript{phe} is more prokaryote-like than eukaryote-like in that it lacks the typical eukaryotic tRNA\textsuperscript{phe} modified bases. However, in the location of specific nucleotides within the sequence, it has both prokaryotic and eukaryotic characteristics. Thus at positions 4 and 26, it has nucleotides characteristic of eukaryotic tRNAs\textsuperscript{phe}, and at positions 20, 44, 45, and 60, it has nucleotides characteristic of prokaryotic tRNAs\textsuperscript{phe}.
Table 10. Positions in the tRNAs\textsuperscript{phe} Which Have Characteristic (34) Eukaryote Cytoplasmic or Prokaryotic Residues.

<table>
<thead>
<tr>
<th>Position Number</th>
<th>Parent Nucleotide\textsuperscript{a}</th>
<th>\begin{tabular}{l} \text{Eukaryote Cytoplasmic} \\	ext{tRNAPhe} \end{tabular}</th>
<th>\begin{tabular}{l} \text{Prokaryote} \\	ext{tRNAPhe} \end{tabular}</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>G</td>
<td></td>
<td>U, C, A/never G</td>
</tr>
<tr>
<td>20</td>
<td>G</td>
<td></td>
<td>U</td>
</tr>
<tr>
<td>26</td>
<td>G</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>44</td>
<td>A or U</td>
<td></td>
<td>G</td>
</tr>
<tr>
<td>45</td>
<td>G</td>
<td></td>
<td>U</td>
</tr>
<tr>
<td>60</td>
<td>C</td>
<td></td>
<td>U</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Disregarding post-transcriptional modifications
The *Euglena* cytoplasmic tRNA\(^{\text{phe}}\) is a typical eukaryotic tRNA\(^{\text{phe}}\) both in having the characteristic modified bases as well as having "eukaryotic" nucleotides at the specific locations. Strikingly, it shares a far greater sequence homology with the beef liver cytoplasmic tRNA\(^{\text{phe}}\) (94.7%) than with the cytoplasmic tRNAs\(^{\text{phe}}\) of wheat germ (78.9%) or yeast (80.3%). Its sequence, therefore, is more "animal-like" than "plant-like." This information and the results of other sequence studies on *Euglena* macromolecules may eventually be used to clarify problems associated with *Euglena* taxonomy. *Euglena* has long been an enigma to biologists because it has both animal-like and plant-like characteristics. Zoologists have put it in the family Euglenidae within the phylum Protozoa at the base of the animal kingdom. Botanists have put it in the family Euglenaceae as a class of algae at the base of the plant kingdom (123,124). The sequencing of protozoan and higher algae tRNAs\(^{\text{phe}}\) would further resolve this problem.

The nucleotide sequence of blue-green algae tRNA\(^{\text{phe}}\) has recently been elucidated in our laboratory (125) to serve as a comparison with *Euglena* chloroplastic tRNA\(^{\text{phe}}\). Blue-green algae are considered to be similar to the "prokaryotic invaders" which incorporated themselves into other cells and eventually became chloroplasts by those who believe that organelles evolved from prokaryotic endosymbionts (63). Fig. 39 (A) shows the structure of blue-green algae tRNA\(^{\text{phe}}\) in cloverleaf form. Both in the absence of the eukaryotic-like modified bases and the presence of specific nucleotides
within its sequence the blue-green algae tRNA\textsuperscript{phe} is a typical prokaryotic tRNA. It shares an 82.9% homology with the \textit{Euglena} chloroplast tRNA\textsuperscript{phe}.

The sequence of a second chloroplastic tRNA\textsuperscript{phe}, that of the bean \textit{Phaseolus vulgaris}, has recently become available (126). Fig. 39 (B) shows that the structure of bean chloroplastic tRNA\textsuperscript{phe} shares a 93.4% sequence homology with \textit{Euglena} chloroplastic tRNA\textsuperscript{phe}. In terms of location of specific nucleotides within its sequence, this higher plant organelle tRNA\textsuperscript{phe} resembles prokaryotic tRNAs\textsuperscript{phe} even more than \textit{Euglena} chloroplastic tRNA\textsuperscript{phe}. For example, at position 26 bean chloroplast tRNA\textsuperscript{phe} has an A residue, whereas \textit{Euglena} chloroplast tRNA\textsuperscript{phe} has a G (see Table 10). Furthermore, bean chloroplast tRNA\textsuperscript{phe} has a greater sequence homology with blue-green algae tRNA\textsuperscript{phe} than \textit{Euglena} chloroplastic tRNA\textsuperscript{phe} does (86.8% vs. 82.9%).

The information presented above is generally consistent with an endosymbiotic origin for chloroplasts. The high sequence homology between the two chloroplast tRNAs from widely divergent organisms suggests that the chloroplasts of bean and \textit{Euglena} share a common ancestor. The fact that both the chloroplast tRNAs\textsuperscript{phe} have a greater sequence homology with blue-green algae tRNA\textsuperscript{phe} (Table 11) than with any other tRNA\textsuperscript{phe} sequenced to date is consistent with the theory that the endosymbionts (63) for chloroplasts were originally blue-green algae. The fact that bean chloroplast tRNA\textsuperscript{phe} is more homologous with blue-green tRNA\textsuperscript{phe} than \textit{Euglena} chloroplast tRNA\textsuperscript{phe} (86.8 vs. 82.9%) could possibly suggest that the invasion of the ancestor of bean and
Fig. 39. Nucleotide sequences of (a) blue-green algae tRNA^phe and (B) bean chloroplast tRNA^phe. In (B) boxed nucleotides show residues which differ from blue-green algae tRNA^phe.
### Table 11. Sequence Homology Between Various Phenylalanine tRNAs (%)

<table>
<thead>
<tr>
<th></th>
<th>Mycop.</th>
<th>E. coli</th>
<th>B. G. algae (Euglena)</th>
<th>Chl. (Bean)</th>
<th>S. pompe</th>
<th>Yeast</th>
<th>Wheat germ</th>
<th>Mammal</th>
<th>Cyto. Euglena</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B. subtilis</strong></td>
<td>82.9</td>
<td>75.8</td>
<td>81.6</td>
<td>71.0</td>
<td>76.3</td>
<td>64.5</td>
<td>60.5</td>
<td>65.8</td>
<td>63.2</td>
</tr>
<tr>
<td><strong>Mycoplasma</strong></td>
<td>72.4</td>
<td>78.9</td>
<td>68.4</td>
<td>69.7</td>
<td>55.3</td>
<td>63.2</td>
<td>64.5</td>
<td>61.8</td>
<td>61.8</td>
</tr>
<tr>
<td><strong>E. coli</strong></td>
<td>84.2</td>
<td>70.0</td>
<td>78.9</td>
<td>56.6</td>
<td>63.2</td>
<td>64.6</td>
<td>71.0</td>
<td>68.4</td>
<td></td>
</tr>
<tr>
<td><strong>B. G. algae</strong></td>
<td>82.9</td>
<td>86.8</td>
<td>59.2</td>
<td>59.2</td>
<td>69.7</td>
<td>69.7</td>
<td>69.7</td>
<td>67.1</td>
<td>67.1</td>
</tr>
<tr>
<td><strong>Chl. (Euglena)</strong></td>
<td>93.4</td>
<td>60.5</td>
<td>61.8</td>
<td>69.7</td>
<td>67.1</td>
<td>67.1</td>
<td>64.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chl. (Bean)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>S. pompe</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>64.5</td>
<td>65.8</td>
<td>72.4</td>
<td>67.1</td>
</tr>
<tr>
<td><strong>Yeast</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>84.2</td>
<td>77.6</td>
</tr>
<tr>
<td><strong>Wheat germ</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>81.6</td>
<td>78.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mammal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>94.7</td>
</tr>
</tbody>
</table>
the ancestor of *Euglena* by prokaryotic endosymbionts occurred at different times after the divergence of the two ancestral lines. On the other hand, we are not sure if the difference in percent homology has any significance. The remarkable similarity between *Euglena* cytoplasmic tRNA\(^{\text{phe}}\) and beef liver cytoplasmic tRNA\(^{\text{phe}}\) indicates *Euglena* and mammals share a common ancestor; perhaps, the adoption of chloroplasts by the animal-like ancestor of *Euglena* is a comparatively recent phenomenon. Cases of endosymbiotic algae are presently known which inhabit animal cells, retain their cell walls, and can in some cases grow independently of their host. One such example of this is *Cyanophora paradoxa*, a blue-green algae (127) which inhabits an animal cell, has lost its wall, and divides along with the host. Therefore, the origin of chloroplasts in evolution is not necessarily a single event.

The fact that the *Euglena* chloroplast tRNA\(^{\text{phe}}\) has two "eukaryotic nucleotides" may not be significant considering the limited amount of tRNAs\(^{\text{phe}}\) which have been sequenced, and these residues may simply reflect mutations.

More evidence for the prokaryotic origin of the chloroplast was recently provided by the sequencing of the initiator tRNAs from the chloroplasts and cytoplasm of *Phaseolus vulgaris* (128). Like prokaryotic initiators, the chloroplast initiator was found to be a tRNA\(^{\text{fmet}}\). Furthermore, it was found to be structurally identical to prokaryotic initiator tRNAs. Thus it lacked a base pair at the top of its acceptor stem, had an A-U base pair in the D stem instead of the
semi-invariant $Y_{11} \cdot U_{24}$ found in all other tRNAs. It also has the residues $T_{54}$, $\psi_{55}$, and $C_{56}$ as expected for a prokaryotic initiator. In contrast, the cytoplasmic initiator was found to be a typical eukaryotic initiator tRNA.

The initiator tRNA$^{f\text{met}}$ of the mitochondria of *Neurospora* (129) has also been sequenced. The mitochondrial initiator, unlike the chloroplast tRNA$^{\text{phe}}$, does not strongly resemble its prokaryotic counterpart. It shares less homology with either prokaryotic or eukaryotic initiator tRNAs than these groups share with each other. It is slightly more homologous with the prokaryotic initiators than with the eukaryotic initiators. However, it contains a base pair at the end of its acceptor stem (a eukaryotic initiator tRNA feature). Uniquely, it has a very high A and U content, and contains only four modified bases. It contains neither the $T_{54}$, $\psi_{55}$, or $C_{56}$ typical of prokaryotic initiator tRNAs or the $A_{54}$, $U_{55}$ typical of eukaryotic initiators. Barnett (62) conjectures that assuming the mitochondrial tRNA$^{f\text{met}}$ has been subjected to the same selective pressures for conservation of sequence as those initiator tRNAs from prokaryotes and eukaryotes, then the mitochondria may have evolved from a primitive organism which may no longer exist. Furthermore, because of the similarities between the chloroplastic tRNAs$^{\text{phe}}$ and prokaryotic tRNAs$^{\text{phe}}$, the development of chloroplasts is a more recent event in the course of evolution.

The fact that *Euglena* cytoplasmic tRNA$^{\text{phe}}$ strongly resembles mammalian tRNA$^{\text{phe}}$, whereas its chloroplastic counterpart is
"prokaryote-like", seems to give support to the endosymbiotic theory of organelle origin. The sequence of *Euglena* cytoplasmic tRNA\(^\text{phe}\) also brought out interesting implications concerning the taxonomy of *Euglena*. The fact that *Euglena* cytoplasmic tRNA\(^\text{phe}\) shares almost 95% sequence homology with mammalian tRNA\(^\text{phe}\), whereas it shares only about 80% homology with tRNAs\(^\text{phe}\) from plant sources seems to indicate a strong "animal-like" character for *Euglena*. The tools of biochemistry are helping to resolve problems which classical biologists have debated for decades.
REFERENCES


34. Gauss, D. H., Gruter, F., Sprinzl, M., Nucleic Acid Res. 6, r1 (1979)


42. Yourno, I., and Kohno, T., Science 175, 650 (1972).


78. Jekowsky, E., Personal Communication.


80. Chia, L. S. Y., Personal Communication.


86. Peacock, A. C., and Dingman, C. W., Biochemistry 6, 1818 (1967).
106. Randerath, K., Personal Communication.


VITA

Charles Kasriel Brum was born in Paris, France in 1950. At the age of three he immigrated with his parents to the United States and settled in New Orleans, Louisiana.

He attended Fortier High School where he graduated in 1968. He attended the University of New Orleans from which he received a Bachelor of Science degree in biology. He entered the Department of Biochemistry at LSU for graduate studies in the fall of 1974.
Candidate: Charles Kasriel Brum

Major Field: Biochemistry

Title of Thesis: The Primary Structure of Euglena gracilis Cytoplasmic Phenylalanine Transfer RNA

Approved:

[Signatures of Major Professor and Chairman, Dean of the Graduate School]

EXAMINING COMMITTEE:

[Signatures of committee members]

Date of Examination:

April 27, 1979