



University  
of Glasgow

<https://theses.gla.ac.uk/>

Theses Digitisation:

<https://www.gla.ac.uk/myglasgow/research/enlighten/theses/digitisation/>

This is a digitised version of the original print thesis.

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study,  
without prior permission or charge

This work cannot be reproduced or quoted extensively from without first  
obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any  
format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author,  
title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>  
[research-enlighten@glasgow.ac.uk](mailto:research-enlighten@glasgow.ac.uk)

Modified Nucleotides in Eukaryotic Ribosomal RNA

by

David Geoffrey Hughes

Thesis presented for the degree of

Doctor of Philosophy,

The University of Glasgow.

September 1976.

ProQuest Number: 10647071

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10647071

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

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

Thesis  
4456  
Copy 2.





## Acknowledgements

I wish to thank the following for assistance during this project:-

Dr. B. E. H. Maden, my supervisor, for encouragements and guidance in the art of scientific investigation.

Professors R. M. S. Smellie and A. R. Williamson for making available the facilities of the Department of Biochemistry.

The staff of the Wellcome Cell Culture Unit for assistance in the growth of tissue culture cells.

Dr. I. D. Hamilton and Mr. C. Minto for assistance in the growth of *Dictyostelium discoidium*.

Dr. Lesley Coggins of the Beatson Institute for Cancer Research for assistance with electron microscopy.

Miss Marie McKeown for assistance in preparation of this manuscript.

Mrs. M. White for skilled typing of the manuscript.

The S. R. C. for a research studentship.

Various friends in the Department of Biochemistry for helpful discussions.

### Abbreviations

r RNA	Ribosomal RNA
r pre RNA	Ribosomal precursor RNA
r DNA	Ribosomal DNA
t RNA	Transfer RNA
m RNA	Messenger RNA
Hn RNA	Heterogeneous nuclear RNA
C	Cytidine
G	Guanosine
A	Adenosine
U	Uridine
ψ	Pseudouridine
RNase	Ribonuclease
DNase	Deoxyribonuclease
SDS	Sodium dodecyl sulphate
EDTA	Ethylene diamine tetra acetic acid
DEAE	Diethylaminoethyl cellulose
cpm	Counts per minute
S	Svedberg unit.

## CONTENTS

	Page
Title	i
Acknowledgements	ii
Abbreviations	iii
Contents	iv
Summary	ix
<u>Chapter 1 Introduction</u>	1
1.1 The Ribosome	1
1.2 Eukaryotic Ribosomal RNA	2
1. General features	2
2. Transcription of rRNA	4
3. Ribosome Assembly	10
4. Secondary Modification of rRNA	11
5. Secondary Structure of rRNA and r pre RNA	16
1.3 Aims of the Project	26
<u>Chapter 2 Materials and Methods</u>	29
Materials	
Chemicals	29
Solutions, Cells and Media	31

	Page
Methods	
2.1 Growth and Radioactive Labelling of Tissue Culture Cells.	34
1. Routine Maintenance of Cells	34
2. Radioactive Labelling	35
2.2 Preparation of RNA from Tissue Culture Cells.	37
1. Cell Harvesting	37
2. Cell Fractionation	37
3. Purification of RNA	39
2.3 Preparation of Ribosomal RNA from Dictyostelium discoidium	42
2.4 RNA Fingerprinting Technique	43
1. Enzymic Digestion of RNA	43
2. 2-Dimensional fractionation	46
2. Autoradiography	48
4. Recovery of Oligonucleotides from Fingerprints	48
2.5 Analysis of Oligonucleotides and RNA	49
1. Complete Hydrolysis of Oligonucleotides and RNA	49
2. Detection and Quantitation of Pseudouridine	51
3. Digestion of Oligonucleotides with Pancreatic RNase	53
4. Complete Digestion with Snake Venom Phosphodiesterase	54

	Page
2.6 Preparation and Properties of Fragments of HeLa 28S rRNA	54
1. Partial Digestion of 28S rRNA with T <sub>1</sub> RNase	54
2. Polyacrylamide Gel Electrophoresis	55
3. Electron Microscopy of Fragment A	56
4. Bisulphite Modification of RNA	57
<u>Chapter 3 The Pseudouridine Acid Content of</u> <u>Eukaryotic rRNA</u>	59
3.1 HeLa cell rRNA	60
1. Base Composition	60
2. Pseudouridine Content	60
3. Pseudouridine and 2'-O-methylation	64
3.2 The Pseudouridine Content of rRNA from other Eukaryotes	64
1. Base Composition	65
2. Pseudouridine Content	66
3. Pseudouridine and 2'-O-methylation: interspecies comparison	68
3.3 Pseudouridine in HeLa Ribosomal Precursor RNA	69
1. Base Composition	69
2. Pseudouridine Content	70

	Page
<u>Chapter 4</u> <u>Fingerprinting Studies on Pseudouridine</u>	73
<u>in HeLa rRNA</u>	
4.1.        Long $T_1$ Fingerprints of HeLa 18 S rRNA	74
4.2 $T_1$ plus Phosphatase Fingerprints of HeLa	
18 S rRNA	78
4.3        Long $T_1$ Fingerprints of HeLa 28S rRNA	81
4.4        Discussion	82
<u>Chapter 5</u> <u>Partial Digestion of 28S rRNA with</u>	85
<u><math>T_1</math> Ribonuclease</u>	
5.1        Purification and Properties of Fragment A	85
1.    Preparation	85
2.    Base Composition	87
3. $T_1$ Ribonuclease Fingerprints of Fragment A	88
4.    Pancreatic Ribonuclease Fingerprints	
of Fragment A	89
5.    Polyacrylamide Gel Electrophoresis of	
Fragment A	90
5.2        Secondary Structure of Fragment A	91
1.    Denaturing Gels	91
2.    Bisulphite Modification	92
3.    Electron Microscopy of HeLa Fragment A	93

	viii. Page
5.3 Modified Nucleotides in Fragment A	95
1. Methylation	95
2. Pseudouridine	98
5.4 Partial Digestion of other rRNA's	99
5.5. Discussion	101
<u>Chapter 6</u> <u>Discussion</u>	107
6.1 Previous Studies on Pseudouridine in HeLa rRNA	107
6.2 Pseudouridine and 2'-O-methylation in rRNA	109
6.3 Modified Nucleotides and RNA structure	112
6.4 Other Possible Methods of Detection of Pseudouridine in rRNA	116
6.5 The Molecular Basis of Secondary Modification of rRNA	120
6.6 The Function of Secondary Modifications in rRNA	122
<u>Appendix</u>	125
References	126

## Summary

Eukaryotic ribosomal RNA contains a considerable number of modified nucleotides. The most common types of modification are 2'-O-methylation and pseudouridylation. This thesis describes studies on pseudouridine in eukaryotic rRNA and its possible relationship to 2'-O-methylation.

A simple two-dimensional chromatographic system was developed for separating Up and  $\Psi$ p from digests of  $^{32}\text{P}$ -labelled RNA. This system was used to quantitate  $\Psi$ p in HeLa rRNA. The number of pseudouridines was found to be very close to the number of 2'-O-methyl groups. HeLa 28S RNA contains some 62 2'-O-methyl groups and 60 pseudouridines whilst 18S RNA contains approximately 38 2'-O-methylations and 37  $\Psi$ p residues. The  $\Psi$ p content of rRNA from mouse L-cells, Xenopus laevis, and Dictyostelium discoidium was also quantitated. The number of pseudouridines and the number of 2'-O-methyl groups was approximately equal in L-cell rRNA and in Xenopus 28S RNA. There was an appreciable excess of pseudouridine over 2'-O-methylation in Xenopus 18S RNA. The  $\Psi$ p content of Dictyostelium rRNA was markedly lower than in vertebrate rRNA's. Data on the number of 2'-O-methyl groups in this species are not accurate but it appears that methylation is also lower than in vertebrates.

The pseudouridine content of HeLa 32S and 45S



r pre RNA's was investigated. Considerable numbers of  $\psi$ p residues were present in both precursors. 32S RNA was found to contain slightly more pseudouridines than 28S RNA. The estimated number of  $\psi$ p residues in 45S RNA (70) was less than the sum of the numbers found in 18S and 28S RNA's. Possible alternative explanations for this are discussed.

Techniques for RNA fingerprinting and sequence analysis were used to investigate the presence of pseudouridine in oligonucleotides from HeLa 18S rRNA. Pseudouridine occurs only at specific sites in the rRNA sequences.

A large RNA fragment was prepared by digestion of HeLa 28S rRNA with  $T_1$  ribonuclease under mild conditions, followed by separation of the products on sucrose gradients. Base composition analysis showed that the fragment had a very high G + C content (ca. 80%). Electrophoretic separation on neutral polyacrylamide gels suggested that preparations contained several fragments of comparable, but not identical size. Denaturation of fragments on 6M urea gels suggested that these various fragments have at least part of their primary sequence in common. Both pancreatic RNase fingerprints and electrophoresis under denaturing conditions demonstrated that the large fragments contain internal nicks. The integrity of the large fragment must, therefore, be maintained after partial digestion by secondary structural

interactions. Low reactivity of the fragment to sodium bisulphite suggested that a high proportion of G-C base pairs is present. Electron microscopy in 80% formamide showed that the fragment contains a high degree of secondary structure which is resistant to denaturation. The presence of modified nucleotides in the fragment was investigated. Measurement of  $^{14}\text{C}$ -methyl radioactivity showed that few methyl groups are present. This was confirmed by fingerprinting analysis which showed only one or two 2'-O-methylated oligonucleotides. Pseudouridine was also found to be relatively less abundant in G-C rich fragments than in 28S rRNA as a whole. Not more than three pseudouridines were present in the fragment, which is several hundred nucleotides long.

## Chapter 1. Introduction

### 1.1 The Ribosome

Ribosomes were first noted as "small particles" during light microscopy of Rous sarcoma virus and a variety of eukaryotic tissues (Claude, 1941). Jeneer and Brachet (1941, 1942) identified RNA as a component of these particles and suggested their involvement in protein synthesis. Confirmation that ribosomes are the site of protein synthesis in the cell has led to extensive research on their structure, function, and genetics. (reviewed by Tissières, 1974).

Ribosomes are found universally in cells of both eukaryotic and prokaryotic origin. They may be divided into two main classes. 80S ribosomes are found in eukaryotic cells both free in the cytoplasm and bound to endoplasmic reticulum. Ribosomes of prokaryotic cells and of eukaryotic mitochondria and chloroplasts sediment at 70S. Both classes consist of two subunits of unequal size whose mass is made up of about one third protein and two thirds RNA. There are, however, considerable differences between 80S and 70S ribosomes in both RNA and protein components. Table 1.1 outlines the major distinguishing features between the ribosomes of mammals and *E. coli*.

Differences between the rRNA's include their size, degree and type of secondary modification, and the presence of a low molecular weight, "5.8 S" RNA in eukaryotes. Sinclair and Brown (1971), using nucleic acid hybridization techniques, have

TABLE 1.1

Comparison of Ribosomal Components from Mammalian Cells and E. coli.

	Small Subunit		Large Subunit	
	Mammalian	E. Coli	Mammalian	E. Coli
Sedimentation coefficient of subunit	40S	30S	60S	50S
Sedimentation coefficient of RNA	18S	16S	28S + 5.8S and 5S	23S 5S
Approx. chain length of RNA (nucleotides)	2,000	1,600	5,000 + 158 5S = 120	3,200 120
Methyl groups per RNA	45	14	68 + 2 55, none	16 0
Type of methylation	Mainly ribose	Mainly base	Mainly ribose	Mainly base
Proteins per subunit	30	21	40	34

Adapted from Maden (1976).

Values for distantly related eukaryotes are closer to those for mammals than E. coli. except that the chain length of 28S RNA is shorter in lower eukaryotes.

shown that there is nucleotide sequence homology between rRNA's from a wide range of eukaryotes. They did not, however, detect homology between rRNA of eukaryotes and prokaryotes.

Similarly, homologies between ribosomal proteins from distantly related eukaryotes are detected by polyacrylamide gel electrophoresis (Delaunay et al, 1973). Comparison of ribosomal proteins of eukaryotes and prokaryotes by immunological techniques has revealed only one case of homology (Wool and Stöffler, 1974).

Despite their significant differences in structure the ribosomes of eukaryotes and prokaryotes perform the same basic function in protein synthesis. It is therefore likely that homologous features will come to light.

This thesis is concerned with studies on the structure of eukaryotic rRNA. Most of this introduction is concerned with aspects of that subject. Prokaryotic rRNA and other related topics will, however, be discussed when they seem relevant.

## 1.2 Eukaryotic Ribosomal RNA

### 1.2.1 General features

#### 1.2.1.1 High Molecular Weight rRNA

As outlined in table 1.1 the eukaryotic ribosome contains one high molecular weight RNA in each subunit. In HeLa cells these species have molecular weights of  $0.70 \times 10^6$  Daltons (small subunit) and  $1.75 \times 10^6$  Daltons (large subunit) and sedimentation coefficients of 18S and 28S respectively

TABLE 1.2

Molecular Weight and G + C content of r RNA

Species	Molecular Weight ( $\times 10^{-6}$ )		% G + C	
	Large Subunit	Small Subunit	Large Subunit	Small Subunit
HeLa	1.75	0.70	67	59
Rat (liver)	1.75	0.70	63	58
Mouse (Liver)	1.71	0.70		
Xenopus (liver)	1.54	0.69	64	53
D. Melanogaster	1.40	0.73	40	41
Saccharomyces	1.30	0.72	48	45
Dictyostelium			43	42
Tetrahymena	1.30	0.69	43 (unfractionated r RNA)	
E. Coli	1.07	0.56	54	54

Adapted from Loening (1968) and  
Attardi & Amaldi (1970).

(Loening, 1968). Within eukaryotes the size of the RNA of the small ribosomal subunit is relatively constant. In contrast, the major RNA of the large subunit increases in size with the evolutionary complexity of the organism (table 1.2). An evolutionary trend is also notable in the chemical composition of rRNA. The percentage G + C content of both high molecular weight species increases on progressing from lower eukaryotes to mammals. In addition to the four main types of nucleotide considerable numbers of modified nucleotides are present in rRNA. The most common modifications are methylations and pseudouridylations.

#### 1.2.1.2 Low Molecular Weight rRNA.

In addition to its two high molecular weight RNA's the eukaryotic ribosome contains two low molecular weight RNA species. Both are associated with the large ribosomal subunit.

5.8S rRNA (originally designated 7S by Weinberg & Penman, 1968 and Pene et al, 1968) is released from 28S rRNA by treatments which disrupt hydrogen bonding (Pene et al, 1968). The molecule is approximately 160 nucleotides long and contains several modified nucleotides (Nazar et al., 1975 a, b; Rubin, 1973; Khan & Maden, 1976a) 5.8S rRNA is methylated at two sites in vertebrates but is unmethylated in yeast.  $\psi$  is present in all the species studied (two residues in vertebrates, one in yeast). Pene et al. (1968) showed that 5.8S rRNA is absent from prokaryotes.

5S RNA is present in the larger subunit of both eukaryotic and prokaryotic ribosomes. The polynucleotide chain is about 120 nucleotides long but appears to lack any methylations or pseudouridylations (Review, Monier, 1974).

### 1.2.2 Transcription of rRNA

#### 1.2.2.1 Ribosomal Genes

The genes coding for eukaryotic rRNA (except 5S RNA) are located in the nucleolus. Brown and Gurdon (1964) showed that mutants of *Xenopus laevis* which do not form nucleoli synthesize no rRNA. It was demonstrated by Wallace and Birnstiel (1966) that these homozygous anucleolate mutants lack rDNA completely. *Drosophila* strains which contain different doses of the nucleolar organizer region were found to contain rDNA in amounts proportional to the number of nucleoli present (Ritossa & Spiegelman, 1965). Localization of rDNA within the nucleolus was confirmed by in situ hybridization experiments (reviewed by Reeder, 1974).

Eukaryotic rRNA genes are present in multiple copies in all the species so far examined (Birnstiel, Chipchase & Spiers, 1971). There appears to be a rough correlation between evolutionary complexity and an increased number of ribosomal genes. An unusual situation is found in *Xenopus laevis* in which the 450 gene copies found in somatic cells (Wallace & Birnstiel, 1966) are amplified several thousand times



in the oocyte nucleus (Brown & Dawid, 1968; Gall, 1968).

Besides being present in multiple copies rRNA cistrons have a high G + C content relative to bulk DNA and can thus be separated from it by virtue of a higher buoyant density in  $\text{CsCl}$  (at least in *Xenopus*)(Wallace & Birnstiel, 1966; Sinclair & Brown, 1971). Because of this pure *Xenopus* rDNA can be prepared. This has provided the most accessible system for genetic analysis in eukaryotes.

The structural organization of rDNA has been established both by biochemical investigations (Birnstiel et al, 1968; Brown & Weber, 1968; Dawid, Brown & Reeder, 1970) and by electron microscopic studies (Wellauer & Dawid, 1974; Wellauer & Reeder, 1975; Wellauer et al., 1974a). The repeating unit consists of the genes for 28S and 18S rRNA separated by transcribed spacer DNA with a further transcribed spacer region external to the 18S gene. Attached to this is an extensive non-transcribed spacer region. Besides the expected variation between species in the gene region there is variation in the size of both transcribed and non-transcribed spacers (see fig. 1.1. a).

#### 1.2.2.2 Transcription

Eukaryotic rRNA is initially transcribed as a large precursor molecule which contains the sequences of 18S, 28S and 5.8S rRNA besides considerable transcribed spacer sections (fig. 1.1. b). In

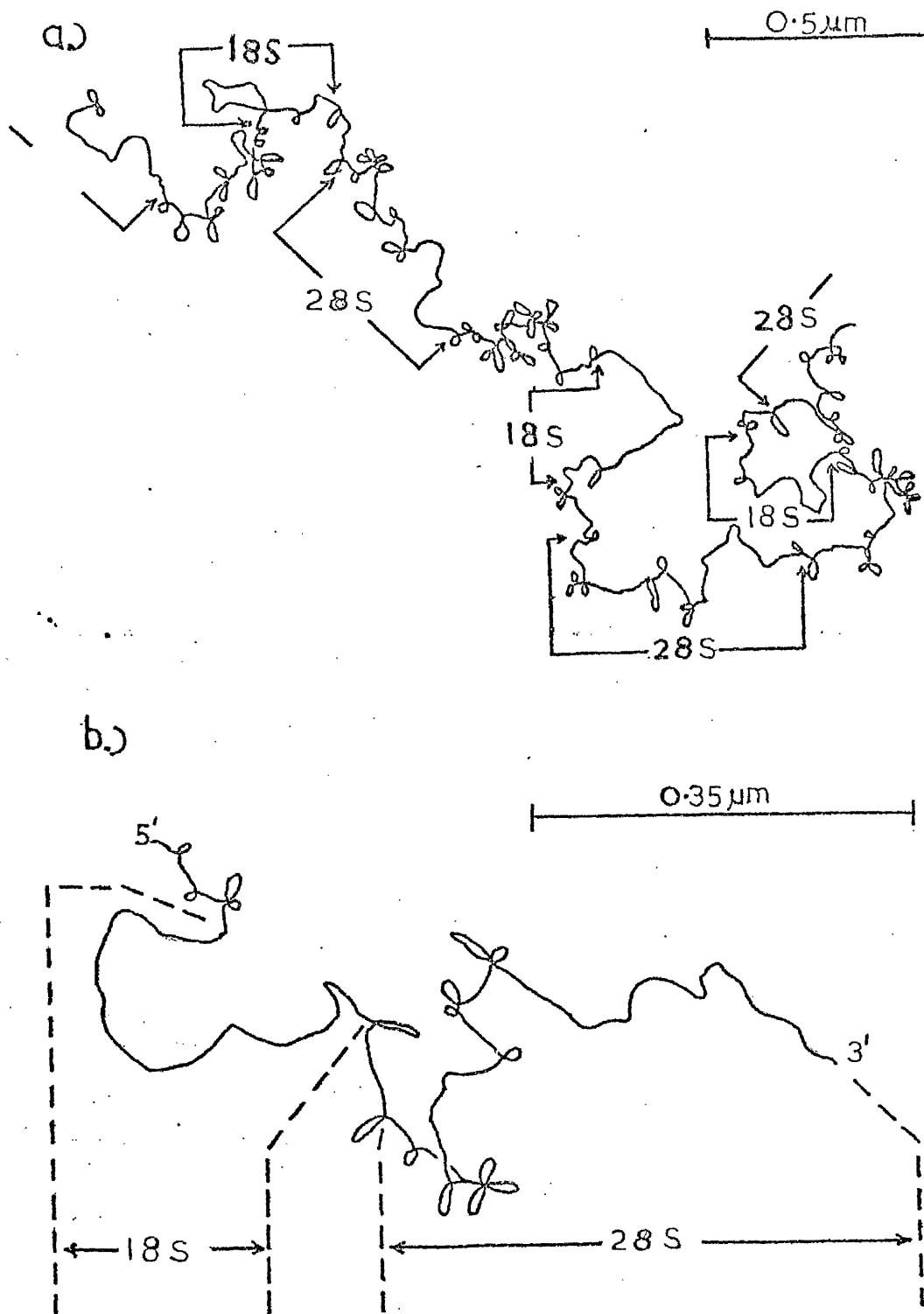


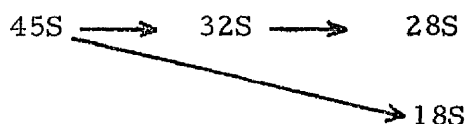
Fig 1.1 Tracings from electron micrographs.

(a) *Xenopus laevis* rDNA

(b) *Xenopus laevis* rRNA

(Adapted from Wellauer & Dawid, 1974)

HeLa cells this molecule contains some 13,000 nucleotides and sediments at 45S. The precursor-product relationship between 45S RNA and mature cytoplasmic rRNA was established by following the kinetics of labelling cellular RNA with radioactive uridine (Scherrer & Darnell, 1962; Scherrer, Latham & Darnell, 1963). The flow of radioactivity was shown to be:-



45S and 32S species are confined to the nucleus. 28S rRNA is detected first in the nucleoplasm and shortly afterwards in the cytoplasm (Penman, 1966; Penman, Smith & Holtzman, 1966). 18S RNA is found only in cytoplasm. If cells are pulse labelled and then treated with Actinomycin D to block further rRNA synthesis radioactivity can be "chased" from 45S into 32S, 18S and later 28S RNA (Girard et al., 1964).

Kinetic experiments with  $^{14}\text{C}$ -Me Methionine as the radioactive species have demonstrated that both 45S and 32S RNA are methylated (Greenberg & Penman, 1966; Zimmerman & Holler, 1967). Vaughan et al. (1967) showed that methylation is essential for ribosome maturation.

Analysis of HeLa RNA on polyacrylamide gels led to the discovery of two short-lived intermediates in the rRNA maturation pathway (41S and 20S RNA), (Weinberg et al., 1967).

Weinberg & Penman (1970) infected HeLa cells with poliovirus causing inhibition of ribosome formation and accumulation of intermediates of the pathway. The level of methylation of these intermediates was analysed and their molecular weights estimated. Results suggested that  $^{41}\text{S}$  RNA contains the same number of methyl groups as 45S RNA, 32S RNA contains the same number of methyl groups as 28S RNA and 20S the same number as 18S RNA. From these experiments the maturation pathway shown in fig. 1.2a was proposed for HeLa cells. Weinberg and Penman also observed low-levels of 36S and 24S RNA species but concluded that they were aberrant cleavage products of 45S rRNA.

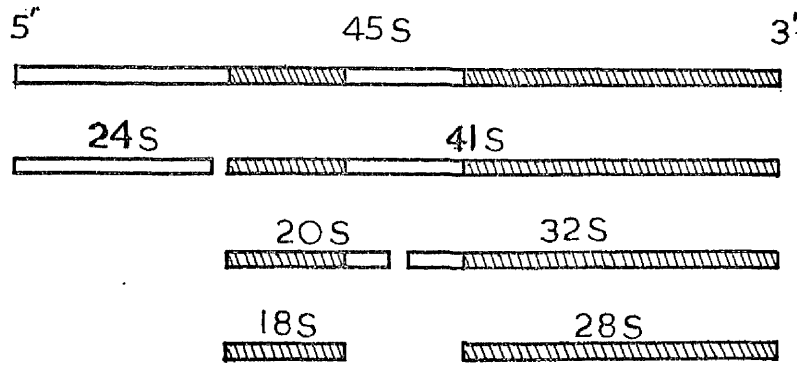
RNA finger-printing studies have conclusively demonstrated the conservation of methyl groups from the precursors to the mature rRNA species; and the combined  $^{14}\text{C}$ -Methyl labelled 18S and 28S RNA's give rise to the same pattern as 45S RNA (Maden, et al., 1972; Maden & Salim, 1974).

In addition to methyl groups HeLa 45S and 32S r pre RNA have been shown to contain a number of pseudouridines (Jeanteur et al., 1968).

Studies on mouse L-cells (Perry & Kelly, 1972) have indicated a processing sequence differing in detail from that for HeLa cells (fig. 1.2b). This scheme includes intermediates of 36S and 24S. A mutant of BHK cells which is temperature sensitive in 28S rRNA production apparently follows the HeLa-type pathway at its permissive temperature

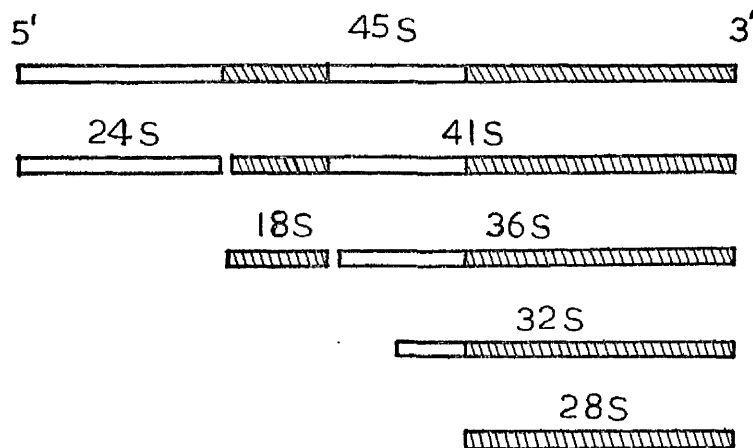
Fig. 1.2 Maturation of 45S rpre RNA.

a.



HeLa -Adapted from Wellauer and Dawid, 1973.

b.



L-cell -Adapted from Wellauer et al, 1974.

(33.5°C) whilst at its non-permissive temperature L-type maturation predominates. It seems that the exact course of rRNA maturation in a given cell type may depend on metabolic conditions (Winicov, 1976).

#### Non-ribosomal sequences (transcribed spacers)

Although estimates of the molecular weight of HeLa r pre RNA and rRNA vary somewhat it is clear that about half the transcribed material is lost during processing (review, Maden, 1971). 45S RNA competes with 28S and 18S RNA for sites in rDNA to the extent expected if 35% and 13% respectively of the precursor molecule contained these ribosomal sequences (Jeanteur & Attardi, 1969). All methyl groups are conserved and hence the removed sequences must be unmethylated (Vaughan et al., 1967; Weinberg et al., 1967; Weinberg & Penman, 1970; Maden & Salim, 1974). Base composition data from two groups (Amaldi & Attardi, 1968; Jeanteur et al., 1968; Willems et al., 1968) purifying r preRNA by different methods show that the G + C content of 45S RNA (70%) is higher than that of either 28S (67%) or 18S (56-58%).

This implies that the transcribed-spacer material has a very high G + C content (about 77%).

#### Polarity of Transcription

The order in which 18S and 28S rRNA's are transcribed in their nucleolar precursor has been in dispute and evidence has been presented for both possible polarities (reviewed in

Dawid and Wellauer, 1976). Partial digestion of 28S rRNA with a supposed 3' exonuclease and electron microscopy of the products led Wellauer and Dawid (1973) to conclude that 28S rRNA is at the 5' end of the 45S molecule. Results from two subsequent studies on the biosynthesis of 45S RNA have strongly suggested the reverse of this polarity (Liau & Hurlbet, 1975; Hackett & Sauerbier, 1975). This conflict has been resolved by two investigations making use of specific sites in rRNA which are cut by Eco RI endonuclease (Dawid & Wellauer, 1976; Reeder et al., 1976). Both investigations confirm the 5'-18S-28S-3' polarity of transcription. The authors of the first of these studies concede that their earlier results may be erroneous due to incomplete characterisation of the exonuclease used.

#### Transcription of 5.8S rRNA

Labelling experiments with radioactive uridine suggested that 28S and 5.8S RNA are derived from a common precursor (Pene et al., 1968). Pene et al (1968) and Udem & Warner (1972) showed that treatments which disrupt hydrogen bonding will release 5.8S RNA from 28S RNA but not from 32S or 45S RNA, implying that if 5.8S RNA is attached to these molecules the linkage is covalent. Maden & Robertson (1974) have shown that oligonucleotide spots characteristic of HeLa 5.8S rRNA are present in fingerprints of 32S r preRNA, confirming the presence of the 5.8S sequence in the precursor. In addition

hybridization experiments using 5.8S rRNA and rDNA indicate that the 5.8S sequence is located between 18S and 28S sequences (Spiers & Birnstiel, 1974). Together these results show that 5.8S rRNA is transcribed as part of 45S r pre RNA and is located in the transcribed-spacer region of the 32S sequence. There is now evidence of the existence of intermediates in the maturation of 5.8S RNA both in yeast and HeLa cells (Trapman et al., 1975; Khan & Maden, 1976b).

### 5S RNA

Although 5S RNA is part of the ribosome its transcription is not genetically linked with the other rRNA species and no larger precursor has yet been identified (reviewed by Monier, 1974).

### 1.2.3 Ribosome Assembly

The assembly of ribosomal subunits from their protein and RNA constituents takes place mainly in the nucleolus. 45S and 32S RNA's can be extracted from nucleoli in ribonucleoprotein particles (RNP's) sedimenting at 80S and 55S respectively (Warner & Soiero, 1967). These particles contain most of the proteins of the large subunit as well as 5S RNA. Attempts to identify proteins of the small subunit in the 80S RNP have been inconclusive (Shepherd & Maden, 1972). This may be due to preferential stripping off of these proteins by the harsh extraction procedure used. In addition to ribosomal proteins nucleolar RNP contains several larger proteins. It appears that these proteins are removed from the RNP before leaving the nucleolus and are



continuously recycled (Kumar & Warner, 1972).

Newly synthesized 40S ribosomal subunits appear in the cytoplasm before 60S subunits because maturation of their RNA is faster. They do not enter functioning polysomes immediately, however. 18S rRNA undergoes several further methylations which are all on bases (Maden & Salim, 1974; Saponara & Enger, 1974). Extra proteins not involved in nuclear assembly are added to the large subunit. In addition certain proteins of both subunits become phosphorylated before the ribosome is fully mature. (Krystosek, et al., 1974).

#### 1.2.4 Secondary Modification of rRNA

Methyl groups and pseudouridine residues have both been identified in rRNA of all organisms studied. Detailed knowledge of these secondary modifications is likely to be important for the understanding of ribosome function.

##### 1.2.4.1 Methylation.

The 28S and 18S rRNA's of HeLa cells contain respectively about 70 and 46 methyl groups (Maden & Salim, 1974). With the exception of five 28S and seven 18S methyl groups which are attached to bases all methylations occur at the 2'-O position of ribose. Sequencing studies on the oligonucleotides containing methyl groups have shown no common primary structural features which could determine the sites of modification (Maden & Salim, 1974).

Khan & Maden (1976) have conducted comparative studies on methylation in vertebrate rRNA's and found a high degree of conservation. Indeed, many distinctive methylation sites are conserved even in the rRNA's of such distantly related eukaryotes as fruitflies (Maden & Tartof, 1974) and yeast (Klootwijk and Planta, 1974).

The rRNA of prokaryotes is also methylated but in this case the numbers of methyl groups are fewer and most are attached to bases (Fellner, 1974). Comparative studies on a wide range of prokaryotic organisms have shown that there is strong conservation of methylation sites in 16S rRNA (Woese et al., 1975).

#### 1.2.4.2 Pseudouridylation

Pseudouridine ( $\Psi_p$ ) was first discovered as a minor component of tRNA (Cohn, 1957; Davis & Allen, 1957). It has subsequently been shown to be almost universally present in this species of molecule (Barrell & Clark, 1974). The presence of  $\Psi_p$  was detected in "microsomal" RNA preparations by Dunn (1959). Confirmation that  $\Psi_p$  is a component of rRNA was obtained by studies on wheat rRNA (Glitz & Decker, 1963; Singh & Lane, 1964) where it is present in relatively large amounts. Amaldi and Attardi (1968) have shown that  $\Psi_p$  is present in HeLa cells at a frequency of approximately 1.1% of nucleotides in 28S rRNA and 1.5% in 18S rRNA. E. coli rRNA, in contrast to eukaryotic rRNA, contains very few  $\Psi_p$  residues. (Dubin & Gunlap, 1967; Nichols & Lane, 1967;

Fellner, 1969; Dahlberg et al., 1975).

Goldwasser and Heinrikson (1966) reviewed the question of whether pseudouridine is itself incorporated into RNA or whether its presence is due to modification of uridine after formation of the polynucleotide chain. They were unable to draw any firm conclusion at that time. However, evidence in favour of the latter possibility is now strong.

Johnson and Söll (1970) synthesized unmodified tRNA in vitro from a DNA template. On incubation in vitro they were able to show an increase in radioactivity in pseudouridine. Studies of eukaryotic tRNA maturation (Mowshowitz, 1970; Burdon, 1975) have shown that mature tRNA contains more  $\psi_p$  than "immature" tRNA of the same size which has been newly formed from pre tRNA. Cortese et al., (1974) have obtained tRNA from mutants of *S. typhimurium* and *E. coli* which lacks two pseudouridines found in tRNA of wild type organisms. This undermodified tRNA becomes modified in the correct positions when incubated with extracts from wT bacteria.

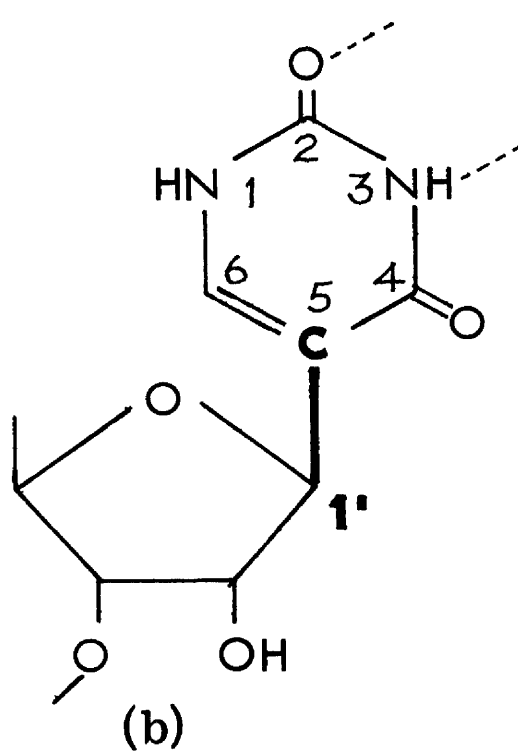
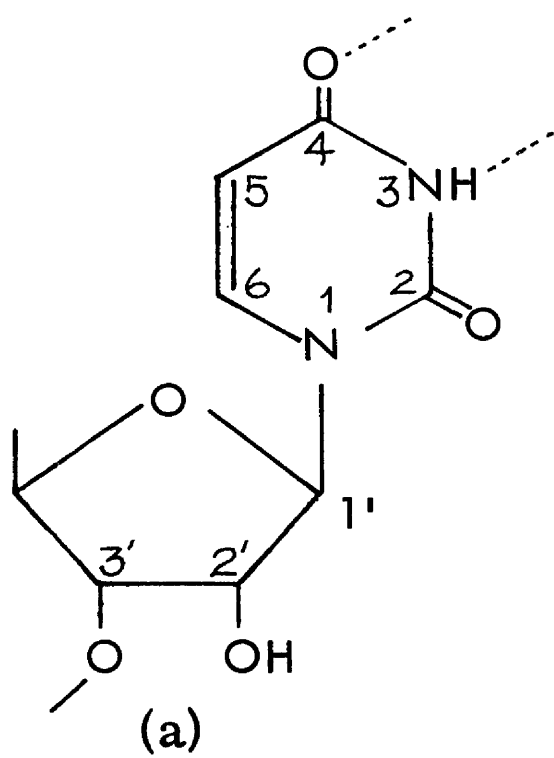
Dubin and Gunlap (1967) found that when *E. coli* cells are treated with chloramphenicol there is a decrease in the proportion of  $\psi_p$  present in rRNA. This suggests that synthesis of some protein necessary for conversion of Up to  $\psi_p$  is blocked.

Amaldi and Attardi (1968) studied the products of ribonuclease digestion of HeLa rRNA using Dowex columns. They determined

Fig. 1.3

The structure of (a) uridine and (b) pseudouridine.

Dashed lines represent hydrogen bonds.



the frequency of occurrence of short sequences containing  $\Psi_p$  relative to that of homologous sequences containing Up. The relative frequency was found to differ markedly between the various sequences studied. This suggests that  $\Psi_p$  is not evenly distributed through the sequence of HeLa rRNA's. Such a finding is consistent with conversion of Up to  $\Psi_p$  at specific sites in the polynucleotide chain.

The structure of naturally occurring pseudouridine has been established as being 5- $\beta$ -D -ribofuranosyluracil (Davis & Allen, 1957; Cohn, 1960; Michelson & Cohn, 1962) as shown in fig. 1.3 b. Formation of  $\Psi_p$  from Up (fig 1.3 a) requires breakage of the covalent bond between ribose C1' and base N1, rotation of the pyrimidine ring about the N3-C6 axis, and formation of a covalent bond between C1' and C5.

The presence of a carbon-carbon bond between ribose and the uracil ring makes  $\Psi_p$  an unusual nucleoside and confers upon it a number of characteristic chemical properties (detailed review by Chambers, 1966). The C-glycosyl bond is very resistant to cleavage and vigorous treatment with hydrochloric acid results in isomerization rather than hydrolysis (Cohn, 1960). C<sub>4</sub><sup>+</sup> has some carbonium ion character. This results in production of a red-brown colour ( $\lambda$  MAX 515-535 nm), characteristic of keto sugars, on heating  $\Psi_p$  with orcinol. Ribose normally gives rise to a blue-green colour in this reaction.

The presence of two dissociable protons in the pyrimidine ring gives rise to the possibility of two different monoanions ( $pK_a$  9.1) as well as a dianion ( $pK_a$  13). At neutral pH the monoanions predominate and the absorption spectrum has  $\lambda$  MAX at 262 nm. In alkaline solutions appreciable amounts of the dianion are present and at pH 12  $\lambda$  MAX is 286 nm. There is no such "bathochromic shift" in the absorption spectrum of uridine since it cannot form a dianion (Cohn, 1960).

Both the orcinol reaction and the bathochromic shift have proved useful in identification of pseudouridine in hydrolysates of various RNA species (Amaldi & Attardi, 1968; Nichols & Lane, 1967).

Lane (1965) and Nichols and Lane (1968), drew attention to a possible correlation between pseudouridine content and 2'-O-methylation in rRNA. Wheat rRNA had been found to contain a considerable number of both types of modified nucleotide (Glitz & Decker, 1963; Singh & Lane, 1964) whereas E. coli rRNA had been found to contain few of either (Dubin & Gunlap, 1967; Nichols & Lane, 1968). The data upon which this suggestion was based were limited both in their scope and accuracy.

Since that time methylation has been studied in detail in several eukaryotic rRNA's. In all cases a relatively large number of 2'-O-methyl groups have been found (Klootwijk & Planta, 1973; Maden & Salim, 1974; Lau et al., (1975); Khan & Maden, 1976). The absolute numbers of 2'-O-methyl groups

have been accurately determined for yeast rRNA (Klootwijk & Planta, 1973), and HeLa cell rRNA (Maden & Salim, 1974). By contrast the  $\psi_p$  contents of these rRNA's have been examined only incidentally, as parts of other studies (Amaldi & Attardi, 1968; Maden & Forbes, 1972; Klootwijk & Planta, 1974). Quantitative determinations have suffered from the difficulty of separating small numbers of  $\psi_p$  residues from much larger quantities of Up in hydrolysates of rRNA. Consequently there has been little advance in the analysis of the possible correlation between  $\psi_p$  content and 2'-O-methylation.

#### 1.2.5 Secondary Structure of rRNA and r pre RNA

##### 1.2.5.1 High Molecular Weight rRNA

(a) Optical Studies. Optical measurements on high molecular weight rRNA in solution suggest that 60-70% of the nucleotides participate in secondary structural interactions (Cox 1966, 1970; McPhie & Gratzner, 1966; Fuller et al., 1967). The main type of interaction is thought to be the base pairing of short stretches of nucleotides in adjacent regions of the polynucleotide chain to form hairpin loops. These loops have differing amounts of double-helical character (Cox, 1970).

More than 90% of the primary sequence of E. coli 16S rRNA has now been determined. This has made possible the formulation of a secondary structure model for this molecule using the method of Tinoco et al., (1971) (Ehresmann et al., 1972; Fellner, 1974). The model contains some 40 hairpin loops involving



base pairing 63% of the nucleotides. This agrees well with previous physical measurements. Studies on the reactivity of 16S rRNA to kethoxal have, however, produced results which are only partially consistent with the simple hairpin-loop model (Noller, 1974). It is probable that tertiary interactions also occur (compare Ladner et al., 1975 on tRNA). Refinements to the proposed model will be necessary as further data on higher-order structure of rRNA become available.

Comparison of hypochromicity profiles of *E. coli* 23S RNA and rabbit reticulocyte 28S rRNA suggests a marked difference in their secondary structures. 23S rRNA shows a monophasic transition suggesting a "melting-out" of double-helical regions of about 55% G + C content. Melting of 28S rRNA is biphasic, apparently due to the presence of two distinct types of double helical regions. These regions have respectively 55% and 78% G + C content (Cox, 1966; Cox et al., 1973). Similar measurements on 45S rRNA of Krebs ascites cells (Hadjilov & Cox, 1973) have shown that it has a secondary structure more related to that of 28S rRNA than 18S rRNA.

The overall G + C content of the helical regions of 45S RNA was found to be higher than that of both 28S and 18S rRNA. From this it was deduced that the non-conserved regions of 45S RNA are extremely G - C rich.

Cox et al., (1976 b) have extended the study of hypochromicity to the A-U rich rRNA from two non-vertebrate

eukaryotes. These authors found that in the 28S rRNA of both *Drosophila melanogaster* and *plasmodium knowlesi* there are base paired regions of high A+U content in place of the G-C rich regions of higher eukaryotes.

Further secondary structural information on 28SrRNA has been gained by digestion with ribonuclease under mild conditions. The nuclease cleaves the polynucleotide chain preferentially at positions of low secondary structure and yields resistant fragments of varying sizes (Gould, 1966, 1967; Wikman et al., 1969; Cox et al., 1973). Characterization of large fragments from 28S rRNA has shown them to have G+C contents of around 80%. Melting profiles suggest a high degree of base pairing. Most of the second phase of melting whole 28S rRNA can be accounted for by such fragments (Wikman et al., 1969; Cox et al., 1973, 1976a).

Comparison of digests of 28S rRNA from a number of species has shown that the size of the G-C rich fragments tends to increase on ascending the evolutionary scale (Gould et al., 1966; Pinder et al., 1969). Cox et al., (1973) found that the second phase of the hypochromicity profile of *Xenopus laevis* 28S rRNA is much less pronounced than that of rabbit reticulocyte 28S rRNA. These results suggest that the G-C rich regions in 28S rRNA of mammals are larger than in lower vertebrates.

Hypochromicity profiles of eukaryotic 18S rRNA from several

species show considerable resemblance to that of *E. coli* 16S rRNA (Cox, 1970). The main type of interaction present is the base-pairing of regions of 50 - 55% G + C content. Melting profiles of eukaryotic 18S rRNA's are however, biphasic. The additional transition is much less pronounced than that found in melting profiles of 28S rRNA and corresponds to the melting-out of much shorter regions of secondary structure (Cox, 1966, 1970; Cox et al., 1976 a, b). These regions have high G + C content in vertebrates and high A + U content in lower eukaryotes.

The small size of the regions of high secondary structure in 18S rRNA is reflected in the results of partial digestion experiments. Mild nuclease digestion produces only fragments of low molecular weight (Gould et al., 1966; Pinder et al., 1969).

The investigations described in this section have involved measurements on isolated rRNA in solution. Studies of hypochromicity, optical rotatory dispersion, and X-ray diffraction suggest that the main features of rRNA are similar in neutral solution and in the ribosomal particle (Amaldi & Attardi, 1970, review; Cox et al., 1976a). Circular dichroism studies have shown that the presence of protein alters the melting temperature of double-helical regions. It does not, however, appear to alter the number of base pairs present (Cox et al., 1973b).

(b) Electron Microscopic Studies. Wellauer & Dawid (1973) have exploited a method of secondary structure mapping of RNA by electron microscopy on preparations partially denatured with formamide. Regions with highly stable secondary structure remain

TABLE 1.3

Size and Conservation of r pre RNA as  
determined by electron microscopy

Animal	Mol. wt. $\pm$ S.D. $\times 10^{-6}$ r pre RNA	% of r pre RNA conserved in r RNA
Rudd	2.64 $\pm$ 0.13	75.4
Xenopus	2.81 $\pm$ 0.16	78.0
Lizard	2.78 $\pm$ 0.14	78.6
Chicken	3.68 $\pm$ 0.23	62.2
Pigeon	3.85 $\pm$ 0.19	61.7
Mouse	4.66 $\pm$ 0.36	51.9 *
Rat	4.51 $\pm$ 0.20	53.4
Cat	4.73 $\pm$ 0.18	50.6
Monkey	4.22 $\pm$ 0.19	56.5
Human	4.71 $\pm$ 0.45	51.4 $\pm$

\* Wellauer et al. (1974)

+ Wellauer & Dawid (1973)

Other figures from Schibler et al. (1975)

intact whilst regions of less stable secondary interactions are single-stranded under these conditions. Investigations of this type on rRNA and rpre RNA have confirmed the maturation pathways of rRNA in HeLa and L-cells. Measurement of the length of various rRNA species when seen under the electron microscope has given a further estimate of their molecular weight.

18S rRNA has no reproducible secondary structural features when spread for electron microscopy in 80% formamide. Major loops do, however, occur both in 28S rRNA and in the transcribed-spacer regions of rpre RNA (Fig 1.1.b) (Wellauer & Dawid, 1973, 1974; Wellauer et al., 1974b). The position of these loops is similar in all species examined but the size of the loops differs between species. rRNA and r pre RNA contain a higher proportion of secondary structure in mammals than in lower vertebrates. Particularly noticeable is the variation in the external transcribed-spacer region. This accounts for much of the variation in size of r pre RNA during vertebrate evolution (Table 1.3) (Schibler et al., 1975).

The major structural features observed upon electron microscopy of rRNA are thought to be hairpin loops based on the formation of long double-helical regions. These loops are stabilized by high G + C content. Measurements of hypochromicity have confirmed that 18S rRNA has little secondary structure in formamide solutions.

Under these conditions 28S rRNA retains secondary structure in regions whose melting properties suggest a very high G + C content (Wellauer & Dawid, 1973, 1974).

Godwin et al., (1974) carried out electron microscopy on preparations of large fragments of L-cell 28S rRNA resistant to digestion by  $T_1$  ribonuclease. This showed that the G - C rich fragments retain a high proportion of their secondary structure even in 80% formamide. A very high proportion of the fragments have branched structures. Measurements of contour length suggested that the branches correspond to 120 - 260 base pairs. However, melting profiles suggest that the base-pairing is imperfect. From their appearance in electron micrographs it appears that the fragments have been cleaved from the major structural loops of whole 28S rRNA. No precise location in the 28S structure has been suggested.

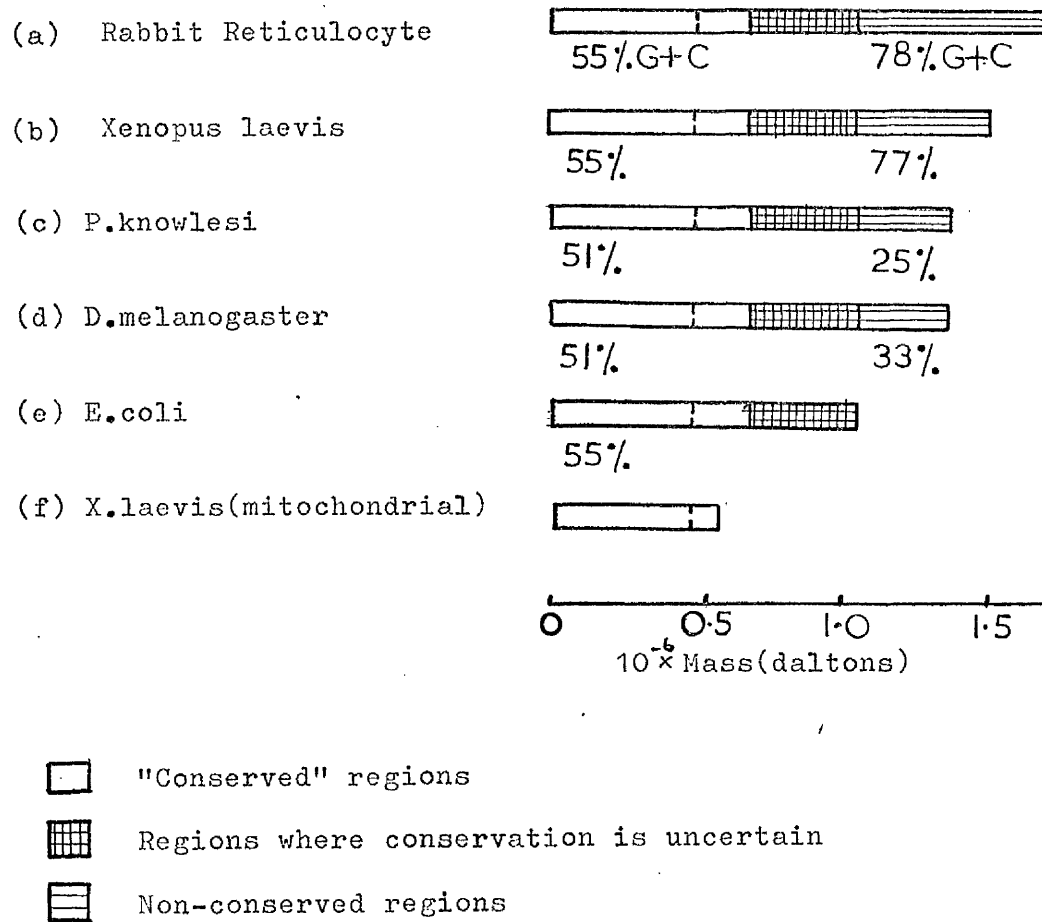
Data on r RNA secondary structure gained from electron microscopy and optical measurements agree well. The predominant feature of 18S secondary structure is the formation of short base paired loops containing roughly equal proportions of A - U and G - C base pairs. Similar features are present in 28S rRNA, but in addition there are regions in 28S rRNA's of atypical secondary structure. These areas have a high content of G - C pairs in vertebrates and a high content of A - U pairs in lower eukaryotes. A - U base pairs are not seen in electron

microscopic studies because they are fully denatured by the conditions used. Any high G - C regions of vertebrate 18S rRNA's are much shorter than in the corresponding 28S rRNA. Because of this they have not been detected as reproducible features in electron microscopy.

Cox et al., (1973, 1976b) have summarized the data on rRNA secondary structure and made suggestions on their evolutionary significance (Fig. 1.4). It is suggested that in both major rRNA species there is a region of 50 - 55% G + C content which is conserved through eukaryote evolution. This agrees with the findings of hybridization studies (Sinclair & Brown, 1971; Birnstiel & Grunstein, 1972). The exact size of this conserved region is uncertain but appears to approach the size of the rRNA's of the smallest ribosomes, the ribosomes of vertebrate mitochondria (Dawid & Chase, 1972). In addition there are non-conserved regions whose base composition varies from 78% G + C in rabbit to 25% G + C in drosophila. The size of the non-conserved areas in 28S rRNA increases through evolution. Non-conserved areas in 18S rRNA are smaller and approximately constant in size.

Extensive conservation of some regions of rRNA throughout eukaryote evolution suggests that they are important in ribosome function. It is likely that these conserved areas of high molecular weight rRNA are involved in those interactions, such

Fig. 1.4 Properties of high molecular weight RNA from the large ribosomal subunit.



(Adapted from Cox et al,1976)



as tRNA binding which are common to all ribosomes. The sequences of the non-conserved regions in rRNA have changed rapidly in evolution. Because of this they are unlikely to be of importance for functions which are common to all ribosomes. Nevertheless the identification of these atypical regions at defined positions provides useful markers in rRNA and r pre RNA. It should be possible to locate the sites of interaction of rRNA with other molecules relative to these features.

One instance of specific interaction between high molecular weight rRNA and other elements of the translational machinery has been noted in prokaryotes. Steitz & Jakes (1975) demonstrated the formation of a complex between the 3'-terminus of E. coli 16S rRNA and the R17 A protein mRNA initiator region. Bacterial 16S rRNA's differ in their 3' terminal sequence. Because of this the degree of base pairing possible with a particular messenger varies between species. This suggests that prokaryotic messengers may be tailored to interact optimally with their own 16S rRNA. The 3'-terminal sequence of eukaryotic 18S rRNA appears to be invariant. A messenger recognition system identical to that in bacteria cannot therefore be formulated for eukaryotes. This may be one reason why it is difficult to express bacterial genes in eukaryotic cells, and vice versa.

Interactions between high molecular weight rRNA and ribosomal proteins are well documented in prokaryotic systems.

(Review by Zimmerman, 1974). Much less information is available on eukaryotes. There are instances in *E. coli* of a single protein binding to sites in 16S rRNA which are widely separated in the primary sequence. Such interactions are not easily explained by the simple hairpin-loop secondary structure as proposed by Ehresmann et al., (1972) (Mackie & Zimmerman, 1975). Immuno-electron microscopic studies have revealed that single ribosomal proteins can, in some instances, bind at two widely separated sites on the ribosome (Tischendorf et al., 1975). This suggests that these proteins probably have extended rather than globular configurations. Complete explanation of RNA-protein interactions in the ribosome will have to take account of the tertiary structure of both types of molecule.

#### 1.2.5.2 Low Molecular Weight rRNA

Neither 5S nor 5.8S rRNA is large enough to be visible by electron microscopy. However, a considerable amount of primary structural information is available on both types of molecule. From these data it has been possible to construct secondary-structural models.

The primary sequences of all the mammalian 5S RNA's studied are identical except for one minor variation (Monier, 1974 review). A number of secondary structure models have been constructed (Monier, 1974; Nishikawa & Takemura, 1974). These

models differ widely in the degree of base-pairing which they propose. Models which maximize base-pairing seem to best explain why some regions of the molecule are attacked by nucleases whilst other regions are resistant.

An intermolecular interaction has been postulated between 5S RNA and tRNA. Prokaryote 5S RNA's have an invariant G-A-A-C sequence which is complementary to G-T- $\psi$ -C found in tRNA (Jordan, 1971). Eukaryotic 5S RNA has G-A-U-C in the position corresponding to G-A-A-C in prokaryotes. Base pairing is possible between this sequence and G-A-U-C which occurs in eukaryotic initiator tRNA in place of G-T- $\psi$ -C (Nishikawa & Takemura, 1974; Simsek et al., 1973). Such interactions could be the site of tRNA recognition by the ribosome.

Complete nucleotide sequences have been determined for 5.8S RNA of rat hepatoma (Nazar et al., 1975a) and yeast (Rubin 1973). In addition, comparison has been made of partial sequence data from a number of other eukaryotic species (Khan & Maden, 1976; Nazar et al., 1975b; Nazar et al., 1976). The nucleotide sequences of mammalian 5.8S rRNA's appears to be homologous at all points except their end groups (Nazar et al., 1976). There is even considerable homology (70%) between the 5.8S RNA's of species as widely separated by evolution as rat and yeast. Secondary structure models which maximize base pairing have been proposed for both these molecules (Nazar et al, 1975b).

Despite the incomplete homology of primary sequence the secondary structure models have very similar overall layout.

### 1.3 Aims of the Project

Eukaryotic rRNA contains considerable numbers of methylations and pseudouridylations. Specific labelling of methyl groups with  $^{14}\text{C}$ -Me Methionine has made this type of modification accessible to investigation. Data are available both on the overall numbers of methyl groups and the primary sequences in which they occur. No specific radioactive label is available for pseudouridine. Less direct methods of detection (see below) have been necessary when studying this type of modification.

The pseudouridine content of RNA's is usually examined by complete hydrolysis of the RNA followed by chromatographic separation of the products. Methods of separating  $\psi$ p from Up include chromatography or ion exchange columns (Glitz & Decker, 1963; Amaldi & Attardi, 1968) or on Whatman paper (using the system of Wyatt, 1951, as reported by Fellner, 1969, or the system of Singh & Lane, 1964), or on thin layers (Nishimura, 1972; Rubin, 1973). The procedure of Glitz & Decker (1963) yielded excellent resolution of  $\psi$ p from Up, but the large number of fractions which were obtained made the method laborious.

The first aim of this project was to develop a system for complete resolution of  $\psi$ p from Up by adaptation of micromethods already in use for RNA sequence analysis. This method involved

alkaline or  $T_2$  ribonuclease digestion of RNA, base composition analysis by electrophoresis at pH 3.5, and two-dimensional paper chromatographic separation of  $\psi$ p from Up using the solvent system of Wyatt (1951). The method was used first to quantitate  $\psi$ p in digests of rRNA and r pre RNA from HeLa cells. Later the same type of analysis was carried out on rRNA from several other eukaryotes. The results obtained were considered in the light of Lane's suggested correlation between  $\psi$ p and 2' -O-Me.

Amaldi and Attardi used column chromatography to investigate short oligonucleotides containing  $\psi$ p in digests of HeLa cell rRNA. Their study made no attempt to locate  $\psi$ p in individual sequences longer than trinucleotides. The techniques of RNA fingerprinting and sequencing analysis can be used to characterize nucleotides produced by ribonuclease digestion of quite small amounts of radioactive RNA. Maden and Salim (1974) used these techniques to investigate methylated sequences in digests of HeLa cell rRNA. The second objective of this project was to apply RNA fingerprinting techniques to the study of  $\psi$ p in 18S rRNA of this cell line. In particular I have attempted to identify  $\psi$ p in long unique oligonucleotides produced by digestion of the RNA with either  $T_1$  ribonuclease or a combination of  $T_1$  ribonuclease and bacterial alkaline phosphatase.

The sites of methylation in rRNA appear to have no primary structural feature in common. It may be that the sites of secondary

modifications are determined by particular secondary structural features. The exact location of secondary modifications within rRNA molecules is unknown.

As discussed above, regions of unusual primary and secondary structure have been detected in eukaryotic rRNA. These regions have high G + C content in vertebrates and high A + U content in lower eukaryotes. Fragments of high G + C content can be isolated from the products of digestion of vertebrate 28S rRNA with ribonucleases under mild conditions.

The third aim of this project was to study the products of  $T_1$  ribonuclease digestion of HeLa-cell 28S rRNA under mild conditions. A large G-C rich fragment was purified on sucrose density gradients. The properties of this fragment were investigated. Particular attention was paid to the level of modified nucleotides present. Experiments were also carried out to investigate the degree of secondary structural interaction in the fragment.

## Chapter 2. Materials and Methods

### Chemicals

All chemicals used were AnalaR reagents from B. D. H. Ltd., Poole, Dorset with the exception of those listed below.

Triton X - 100

Tween 80

Koch - Light Laboratories Ltd.,

Toluene

Colnbrooke, England.

2, 5 diphenyloxazole (PPO)

---

Calf Serum

Foetal Calf Serum

Bio-Cult Laboratories Ltd.,

Amino Acids

Paisley, Scotland.

Vitamins

---

Penicillin

Glaxo Pharmaceuticals, London.

Streptomycin

---

Trypsin

Difco Laboratories, Michigan,  
U. S. A.

Bactopeptone

Yeast extract

---

Kodirex KD 54T (35 x 43) cm

X-ray film

DX - 80 Developer

Kodak Ltd.,

FX-40 X-ray liquid fixer

Cellulose Acetate  
Electrophoresis strips

Oxoid Ltd., London

Whatman DE81 paper

(46 cm x 50 m)

Whatman No. 52 & 3 mm paper

H. Reeve - Angel & Co. Ltd.,  
London.

Cellogel Strips (2.5 x 95 cm)

T<sub>1</sub> & T<sub>2</sub> Ribonucleases

(crystalline) (made by

(Sankyo Co. Ltd., Japan)

Calbiochem Ltd.,

Actinomycin D (200 µg/vial)

Hereford, England.

Pancreatic Ribonuclease

(chromatographically  
homogeneous)

Bacterial Alkaline Phosphatase

Worthington Biochemical

(electrophoretically purified)

Corporation via Cambrian

Snake venom phosphodiesterase

Chemicals Ltd., London.

<sup>32</sup>P-orthophosphate (<sup>32</sup>Pi;

10 mCi/m mol) carrier free

(in solution in dil. HCl.

Radiochemical Centre,

L-(methyl-<sup>14</sup>C) methionine

Amersham.

( 50 mCi/m mol)

(supplied as a solid)



Bromocresol purple	May & Baker Ltd., Dagenham, England.
--------------------	---

---

trypan red

xylene cyanol FF	G. T. Gurr, Ltd.,
eosine	London

fluorescein

---

Deoxycholic Acid (Sodium Salt)

Deoxyribonuclease (RNase free)

Cytochrome C Type VI

Sigma Chemical Co. Ltd.,  
London

### Solutions

#### Balanced Salts Solution (BSS: Earle, 1943)

BSS contained 0.116 M NaCl, 5.4 mM KCl, 1 mM MgSO<sub>4</sub>,  
1 mM Na H<sub>2</sub> PO<sub>4</sub>, 1.8 mM CaCl<sub>2</sub> and 0.002% (W/U) phenol  
red. The pH of the solution was adjusted to 7.0 with 5.6%  
(W/V) Na HCO<sub>3</sub>.

#### Trypsin/Versene

Versene 4 parts, trypsin/citrate 1 part (W/U)

Versene 0.6 mM EDTA in PBS(A).

Phosphate Buffered Saline (A) (PBS(A)) 0.17 M NaCl, 3.4 mM KCl,  
10 mM Na<sub>2</sub> H PO<sub>4</sub> and 2.4 mM KH<sub>2</sub> PO<sub>4</sub>.

#### Trypsin/Citrate

0.25% (W/U) trypsin, 10.5 mM NaCl, 1.0 mM sodium citrate and  
0.002% phenol red adjusted to pH 7.8 with NaOH.

RSB (Reticulocyte Salt Buffer)

0.01M NaCl, 0.0015 M  $\text{MgCl}_2$ , 0.01 M tris-HCl, pH 7.4.

HSB (High Salt Buffer)

0.7 M NaCl, 0.07 M  $\text{MgCl}_2$ , 0.01M tris -HCl, pH 7.4.

SUP (Sodium dodecyl sulphate-Urea-Phosphate buffer)

0.5% (W/V) SDS, 0.5 M Urea, 0.01 M Na phosphate, pH 7.0

LETS (Lithium-EDTA-tris - Sodium dodecyl sulphate buffer)

0.1 M Li Cl, 0.01 M EDTA, 0.01 M tris - HCl, 0.2% SDS, pH 7.4.

Paper electrophoresis dye mixture

1% xylene cyanol F.F. 2% orange G; 1% acid fuschin.

PAGE dye/sucrose mixture

trypan red, 5g/l; xylene cyanol FF, 2g/l; eosine, 10 g/l;

bromophenol blue, 2g/l; bromocresol purple, 2 g/l; fluorescein,

10 g/l; sucrose 500 g/l.

Scintillants

Scintillation counting of non-aqueous samples was carried out in scintillant prepared by dissolving 5 g of 2,5 diphenyl oxazole (PPO) in one litre of toluene.

Counting of aqueous samples was carried out in scintillant prepared by dissolving 5g PPO and 0.5g p-Bis (0-methyl styryl) benzene (Bis-mSB) in 650 ml toluene and 350 ml triton X-100.

Cells and Growth MediaCells

HeLa cells (Gay et al., 1952), L929 cells (Sanford et al., 1948)

and *Xenopus laevis* kidney cells (originally a gift from Dr. K. Jones) growing in monolayer culture and axenic *Amoebae* of *dictyostelium discoidium* grown in suspension culture (Watts & Ashworth, 1970) were used in this study.

### Media

(1) Monolayer cells were grown in slightly modified Eagle's MEM (Eagle, 1959) with the addition of serum to 10% / V/V) (Calf serum for HeLa and L929 cells, foetal calf serum for *Xenopus laevis* cells), penicillin (100 units/ml) and streptomycin sulphate (100 µg/ml). *Xenopus laevis* cells also required the addition of non-essential amino acids to 1%. The constituents of the medium are shown in table 2.1.

For labelling with  $^{32}\text{Pi}$ , cells were grown in medium with  $1/9\text{th}$  ( $1.0 \times 10^{-4}\text{M}$ ) the concentration of phosphate of the medium described above.

When cells were labelled with  $^{14}\text{C}$ -methionine the regular medium was replaced by "minus methionine" Eagle's medium containing 5% dialyzed serum, and  $2 \times 10^{-5}\text{M}$ -adenosine and guanosine and  $10^{-2}\text{M}$  sodium formate to block equilibration of methyl label with the "one-carbon pool" and purine biosynthetic pathway (Winocour et al., 1965; Maden & Salim, 1974).

(2) *Dictyostelium discoidium* cells were grown routinely in axenic liquid medium (Watts & Ashworth, 1970) whose constituents are shown in table 2.2. a. For labelling with  $^{32}\text{Pi}$

TABLE 2.1

## Eagle's Minimum Essential Medium (modified)

<u>Amino Acids</u>	mg/L	<u>Vitamins</u>	mg/L
L-arginine HCl	126.4	D-Ca pantothenate	2.0
L-cystine	24.0	Choline chloride	2.0
L-glutamine	292.3	Folic acid	2.0
L-histidine HCl	41.9	i - Inositol	4.0
L-isoleucine	52.5	Nicotinamide	2.0
L-leucine	52.5	Pyridoxal HCl	2.0
L-lysine HCl	73.1	Riboflavine	0.2
L-methionine	14.9	Thiamine HCl	2.0
L-phenylalanine	33.0		
L-threonine	47.6		
L-tryptophan	10.2		
L-tyrosine	36.2		
L-valine	46.9		

INORGANIC SALTS AND OTHER COMPONENTS

CaCl <sub>2</sub>	200	Penicillin	10 <sup>5</sup> units
D-glucose	4 500	Streptomycin	10 <sup>5</sup> µg
Mg SO <sub>4</sub>	98		
KCl	400		
NaCl	6800		
Na H <sub>2</sub> PO <sub>4</sub> 2H <sub>2</sub> O	140		
Phenol Red (Na)	10		

TABLE 2.2

(a) Axenic Liquid Medium

1 litre contains:-

Bactopeptone	14.3g
Yeast extract	6.16g
Glucose	15.4g
$\text{Na}_2\text{HPO}_4$	0.52 g
$\text{KH}_2\text{PO}_4$	0.48 g

pH 6.7

(b) MES HL 5 Medium

1 litre contains:-

Bactopeptone	10g
Yeast extract	5g

Morpholine ethanesulphonic acid  
(MES) 1.1 g

pH 6.4

the cells were grown in low-phosphate MES HL 5 medium (table 2.2.b) as described by Firtel and Lodish (1973).

## Methods

### 2.1. Growth and Radioactive Labelling of Tissue Culture cells.

#### 2.1.1 Routine Maintenance of Cells

(a) HeLa and L929 cells were maintained routinely as monolayer cultures in rotating "burler" bottles (80 oz. clear-glass Winchester) according to the method of House & Wildy (1965).  $25 \times 10^6$  cells were seeded into 180 ml of medium and grown at 37°C for 3-4 days in a 5% CO<sub>2</sub> atmosphere. When the cells were at a density of 150- 200  $\times 10^6$  per burler they were harvested with trypsin/versene (see below).

(b) *Xenopus laevis* cells were maintained as monolayers in Roux bottles.  $5 \times 10^6$  cells were seeded into 50 ml of medium and grown at 25°C in a 5% CO<sub>2</sub> atmosphere. After 3-4 days the cells were harvested when at a density of 25-30  $\times 10^6$  cells per bottle.

For serial passaging cells were counted on a haemocytometer then seeded and grown as already described. Cell cultures were maintained and contamination checks were carried out by the staff of the Wellcome Cell Culture Unit.

#### Contamination checks

All sterile media and passaged cells were checked regularly for bacterial, fungal or PPLO infection as follows:-

#### Bacterial contamination

Aliquots were grown on blood agar plates and brain-heart

infusion broth at 37°C. Results were considered negative if no growth was seen within 7 days.

#### Fungal contamination

Aliquots were added to Sabouraud's medium and grown at 32°C. Again, no growth after 7 days was assumed to indicate the absence of fungal contamination.

#### PPLO infection

PPLO agar plates were seeded with passaged cells by piercing the agar surface with a charged Pasteur pipette. The plates were grown in an atmosphere of 5% CO<sub>2</sub> in N<sub>2</sub> at 37°C for 7 days and examined microscopically for the characteristic "fried egg" appearance of PPLO colonies.

#### 2.1.2 Radioactive labelling

Labelling cells with <sup>32</sup>PO<sub>4</sub>.

(a) HeLa and L929 cells.

To prepare <sup>32</sup>PO<sub>4</sub> labelled RNA 30 x 10<sup>6</sup> cells were seeded into a burler and grown for 48 hours at 37°C as previously described. After this time the medium was poured off and was replaced with 50 ml of medium containing 1/10th the normal concentration of phosphate. Lowering the amount of unlabelled phosphate present during labelling increased the uptake of radioactive phosphate into the cells.

10 mCi of <sup>32</sup>Pi was added to each burler and the cells were

allowed to grow at 37°C for a further 16-18 hours before harvesting.

(b) *Xenopus laevis* cells.

Roux bottles were seeded at  $6 \times 10^6$  cells / bottle and grown for 48 hours at 25°C. Regular medium was then replaced with 15 ml/bottle of "low phosphate" medium and 3-4 mCi of  $^{32}\text{P}$  were added. The cells were grown for a further 16-18 hours before harvesting.

For the preparation of pure 32S nucleolar RNA actinomycin D (40 µg/ml) was added 30 mins before harvesting. (Robertson & Maden, (1973). This inhibits the synthesis of new 45S RNA and allows the existing 45S species to be processed to smaller RNA species, leaving none to contaminate 32S RNA preparations.

Labelling of HeLa cells with  $^{14}\text{CH}_3$  - Methionine

For the preparation of  $^{14}\text{C}$ -methyl labelled RNA  $5 \times 10^6$  cells were seeded into Roux bottles and allowed to grow at 37°C for 24 hours. The normal medium was then removed and the cells were washed with 25 ml of "Met-minus" medium. This was poured off and a further 50 ml of the same medium was added. 125 µCi  $^{14}\text{CH}_3$ -methionine was then added to each bottle and the cells were grown at 37°C for two days more. After this time cultures were harvested with trypsin/versene.



## 2.2. Preparation of RNA from Tissue Culture cells.

Quantities of reagents given in this section refer to operations on 1 burler containing  $10^8$  cells. Where different numbers of cells were involved these quantities were altered proportionately.

### 2.2.1

#### Cell Harvesting

Growth medium was poured off and the cells were washed twice with 20 ml of trypsin/versene which had been prewarmed to 37°C. After the second washing a few ml of trypsin/versene was left in the bottle and the cells were incubated for a few minutes at 37°C. During this time the cells were seen to dislodge from the glass. When all the cells had become detached they were taken up in 40 ml of ice-cold balanced salts solution containing 10% (U/U) serum, to neutralize tryptic activity, and transferred to a 50 ml tube on ice. The cells were pelleted by centrifugation at 2,000 rpm for 2 mins in an MSE Mistral 4L Centrifuge at 4°C. The cell pellet was washed twice by resuspension in 25 ml balance salts solution followed by centrifugation as before.

### 2.2.2 Cell Fractionation

Cells were fractionated into cytoplasm, nucleoli and nucleoplasm by the method of Penman (1969). All operations were carried out at 4°C.

#### (a) Separation of nuclei and cytoplasm

The washed cell pellet was thoroughly drained, vigorously resuspended in 4 ml hypotonic RSB and the cells allowed to swell for 10-15 mins. The cell suspension was homogenized with 15 strokes of a stainless-steel Dounce homogenizer (clearance 0.003"). The homogenate was transferred to 15 ml Corex test tubes and centrifuged at 2,000 r.p.m. for 2 mins. After centrifugation the supernatant was carefully removed from the pelleted material using a Pasteur pipette. This pelleted material was washed by resuspension in a further 2 ml RSB and centrifuged as before.

Supernatant from this washing was removed and added to that already collected. This pooled supernatant was now treated as being cytoplasm and pelleted material was treated as being nuclei.

(b) Separation of nucleoli from nucleoplasm

The nuclear pellet was washed once with 2 ml RSB and resuspended thoroughly in 4 ml RSB. 0.6 ml Magic (2 parts 1% Tween 80, 1 part 10% Sodium deoxycholate) was added and the test tube was vortexed for 15 seconds to ensure thorough mixing. The preparation was centrifuged at 2,000 r.p.m. for 2 mins. and the resulting supernatant discarded.

This procedure strips off the outer nuclear membrane along with any remaining cytoplasm and also removes any residual unbroken cells.

The pellet was resuspended in 2 ml HSB containing 100 µg

DNAase. This action caused the nuclei to rupture resulting in a marked increase in the viscosity of the preparation. The mixture was heated at 37°C whilst being pipetted repeatedly with a Pasteur pipette. After 1-2 mins. the viscosity of the preparation decreased and all visible clumps of material were dispersed. The preparation was rapidly cooled on ice and layered onto a 16 ml, 15-30% Sucrose-HSB gradient in the small bucket of a Beckman SW 27 rotor. Gradients were spun at 17000 r.p.m. for 20 mins. at 4°C. After centrifugation a small pellet of nucleoli remained at the bottom of the gradient. The gradient supernatant was poured off and the pellet was first drained and then carefully washed with 2 ml RSB, taking care not to disturb the precipitated material.

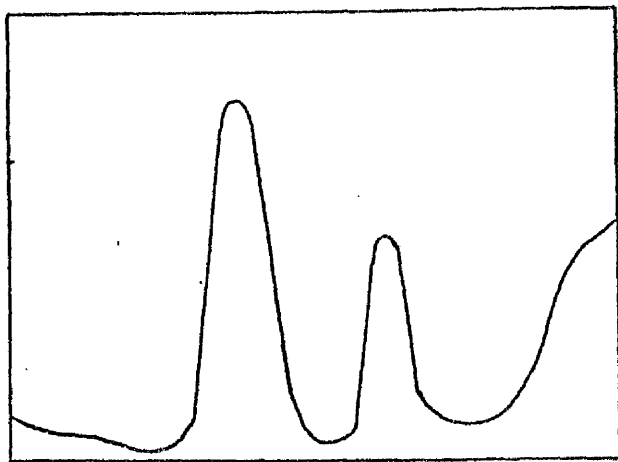
### 2.2.3 Purification of RNA

#### (a) Cytoplasmic RNA.

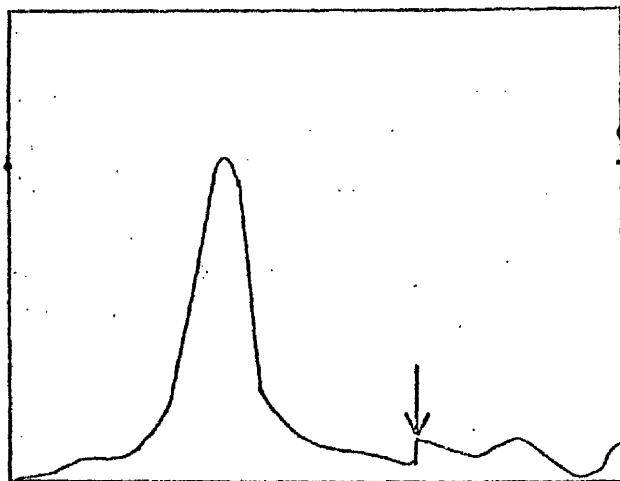
The pooled cytoplasmic supernatant was brought to room temperature and made to 0.5% with SDS. An equal volume of water saturated phenol was added and the preparation was vortexed intermittently for 5 mins. Following this the mixture was separated by centrifugation for 20 mins. at 3500 r.p.m. at 20°C. The aqueous layer was removed and made to 0.2 M in NaCl.  $2\frac{1}{2}$  volumes of cold ethanol was added and the preparation was placed at -20°C to allow precipitation of RNA.

After 4 hours samples were centrifuged for 30 mins. at 3500 r.p.m. and the ethanol was poured off. Precipitated

a.



b.



c.

O.D.  
260

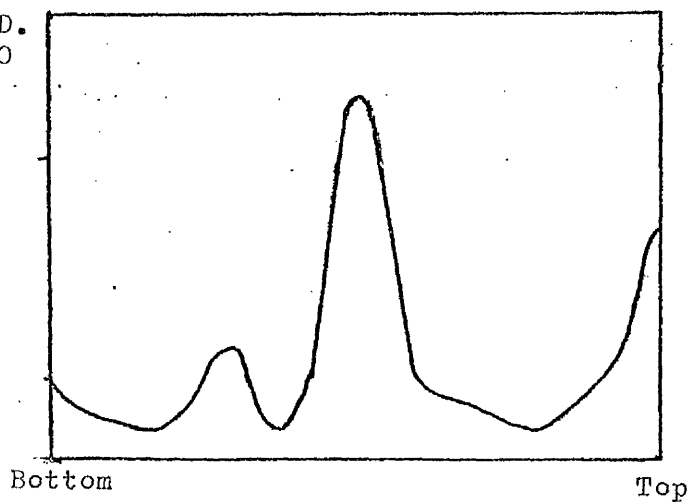


Fig. 2.1 O.D. Profiles of rRNA after sucrose gradient centrifugation.

- (a) Cytoplasmic RNA, 10-25% sucrose-LETS.
- (b) Heat treated 28S RNA, 10-25% sucrose-LETS.
- (c) Nucleolar RNA, 15-30% sucrose-LETS.

RNA was drained, dissolved in 1 ml LETS and layered onto a 37 ml, 10- 25% Sucrose - LETS gradient in the large buckets of the SW 27 rotor. Gradients were centrifuged at 22,000 r.p.m. for 16 hours at 20°C in a Beckman L250 ultracentrifuge. They were then fractionated by pumping through the  $\frac{1}{2}$  cm flow cell of a Gilford recording spectrophotometer and monitoring the optical density at 260 n.m. (see fig. 2.1 a).

Fractions corresponding to 28S and 18S rRNA were collected and precipitated twice with ethanol to remove sucrose and SDS. After the second precipitation RNA was dissolved in water and its specific activity determined as follows:- RNA concentration was determined by measurement of optical density at 260 nm assuming the relationship  $E_{1\text{ cm}}^{1\text{ mg/ml}} = 25$ , and radioactivity was assayed by liquid scintillation counting of 10  $\lambda$  aliquots in triton/toluene scintillant. RNA samples were stored at -20°C until required for use.

$10^8$  cells generally yielded about 500  $\mu\text{g}$  of 28S RNA and 250  $\mu\text{g}$  18S RNA each containing  $5 - 10 \times 10^4$  cpm/ $\mu\text{g}$ .

#### (b) Nucleolar RNA

Nucleoli were resuspended in 1 ml. SUP (Soeiro et al 1968) by repeatedly pipetting with a Pasteur pipette and agitation with a glass rod. This process took 15-30 mins. Nucleolar RNA was extracted with water saturated phenol as described for cytoplasmic RNA and  $2\frac{1}{2}$  volumes of ethanol was added. After precipitation at

-20°C for several hours RNA was pelleted by centrifugation at 3,500 r.p.m. for 30 mins. at 4°C. RNA was drained of ethanol, dissolved in LETS and layered onto 15-30% sucrose -LETS gradients. Gradients were centrifuged at 22,000 r.p.m. for 16 hours at 20°C and fractionated as previously described (see fig. 2.1 c). Fractions corresponding to 45S and 32S were collected and precipitated twice with ethanol. Specific actificity was then measured and samples stored at -20°C.

5. 8S RNA was prepared from 28S rRNA (Maden & Robertson 1974).

After the first ethanol precipitation 28S rRNA was dissolved in 1.0 ml LETS and heated at 60°C for 5 mins with frequent agitation. The preparation was rapidly cooled on ice, layered onto a 37 ml 10 - 25% sucrose - LETS gradient and centrifuged at 22,000 r.p.m. for 16 hours at 20°C. Gradients were fractionated as before and peaks corresponding to 28S and 5.8S were collected (fig. 2.1 b). Carrier RNA was added to the 5.8 S sample. RNAs were precipitated twice with ethanol. Specific activity was assayed and RNA stored at -20°C until required for use.

(c) Preparation of Hn RNA (Fraser et al., 1973)

For preparation of Hn RNA actinomycin D (0.04 µg/ml) was added to the cells 30 minutes before the addition of  $^{32}\text{PO}_4$ . This concentration of actinomycin selectively suppresses rRNA synthesis without apparent effect on Hn or messenger RNA synthesis. The cells were harvested after 3 hours incubation with

$^{32}\text{PO}_4$  and nuclei prepared from them. Nucleoli and nucleoplasm were separated on HSB gradients as described above. RNA was extracted from the supernatant of HSB gradients (nucleoplasm) using water saturated phenol. Nuclear RNA was fractionated by centrifugation on 15-30% Sucrose Lets gradients at 16,000 r.p.m. for 16 hours. Fractions sedimenting faster than 32S were pooled as Hn RNA and precipitated with ethanol.

### 2.3 Preparation of Ribosomal RNA from Dictyostelium discoidium,

Axenically growing amoebae of *Dictyostelium discoidium* were maintained in suspension culture in axenic liquid medium. For radioactive labelling, cells of a third passage were grown to a density of about  $4 \times 10^6$ /ml and then harvested by centrifugation at 5,000 r.p.m. for 15 mins. at 4°C. The cell pellet (ca  $4 \times 10^8$  cells) was washed with 30 ml of low-phosphate MES HL5 medium and then resuspended in 100 ml of the same. The culture was placed on a shaker table at 21°C for 30 mins. after which time 10 mCi  $^{32}\text{PO}_4$  was added. Growth was allowed to continue for 18 hours by which time cells were at a density of approximately  $1.2 \times 10^8$ /ml. Cells were harvested by centrifugation and washed once in 30 ml of MES HL 5. The pellet was resuspended in 6 ml RSB containing 75  $\mu$ l NPT 12 detergent. The preparation was vortexed for 30-45 secs. to allow cell lysis and then centrifuged at 12,000 r.p.m. for 15 mins. The supernatant was removed and was made to 1% in SDS. A room temperature phenol extraction was carried out as

already described. After centrifugation of the first phenol extraction a large amount of flocculating material was present at the interface. The lower (phenol) layer was removed, fresh phenol added and the extraction repeated. The interface after this second extraction contained little flocculating material and the upper (aqueous) layer was removed. RNA was precipitated from this layer with ethanol and fractionated on sucrose gradients as for monolayer cultured cells.

#### 2.4. RNA fingerprinting technique

The techniques of RNA fingerprinting and sequence analysis have been comprehensively described by Sanger et al (1965), Brownlee and Sanger (1967) and Brownlee (1972).

##### 2.4.1 Enzymic Digestion of RNA

The type of enzymes used for RNA digestion depended upon the purpose for which each individual two dimensional separation was required.

##### (a) T<sub>1</sub> ribonuclease and T<sub>1</sub> ribonuclease plus alkaline phosphatase

Digestion of RNA with T<sub>1</sub> ribonuclease yields oligonucleotides with a guanylic acid at the 3' terminus. Sanger et al (1965) described the oligonucleotide "graticules" obtained by this type of fractionation (Fig. 2.2. a & b). The mobility of oligonucleotides is influenced particularly by the number of uridylic acid residues present. Increasing content of uridylic acid causes reduced mobility in the second dimension of the



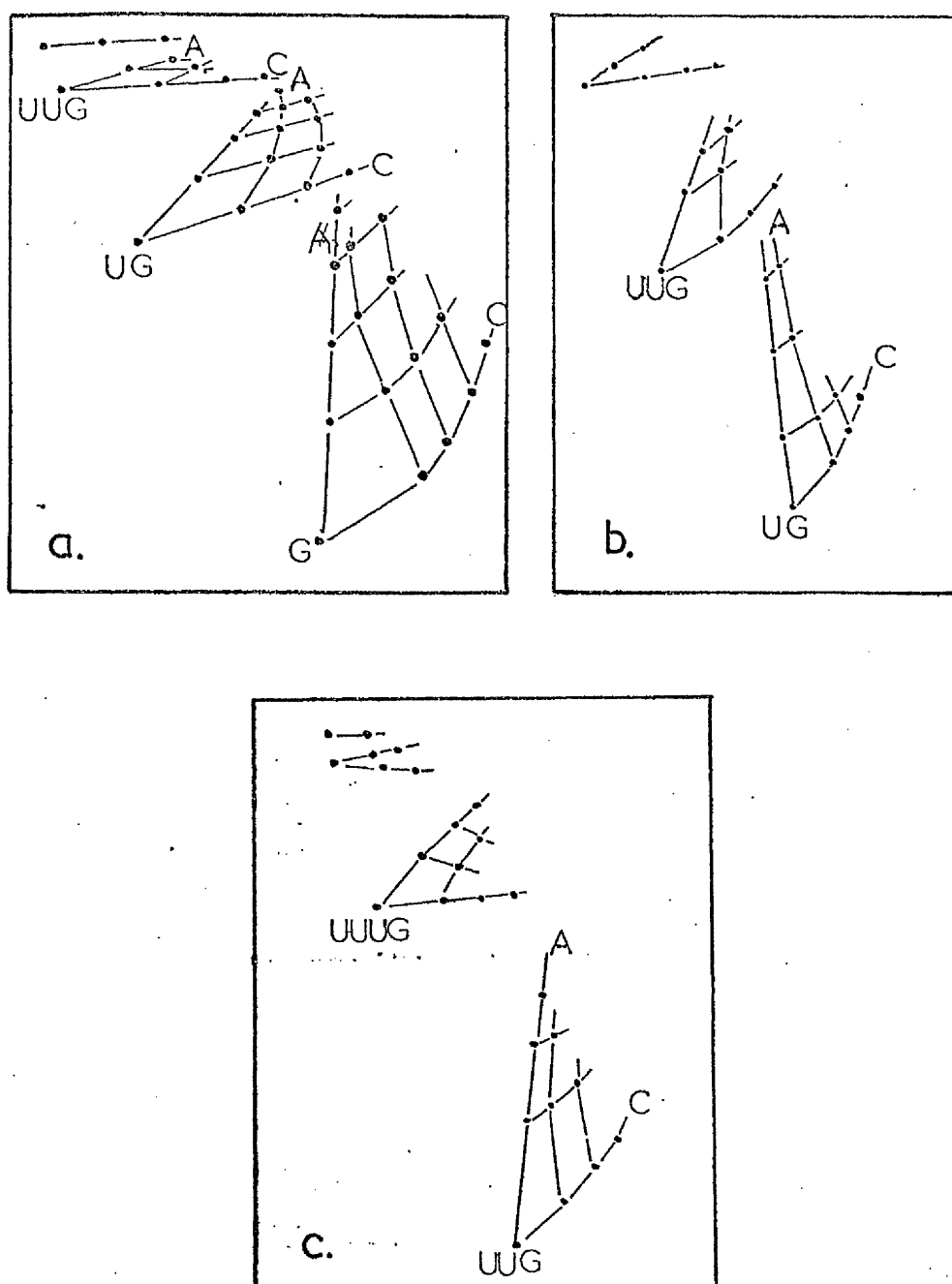


Fig. 2.2 Two-dimensional fractionation of oligonucleotides after  $T_1$  RNase digestion.

(a) Standard  $T_1$  fingerprint.

(b) Long  $T_1$  fingerprint.

(c)  $T_1$  plus phosphatase fingerprint.

(Adapted from Brownlee, 1972)

fractionation system. All oligonucleotides in the same graticule contain the same number of uridylic acid residues.

Electrophoresis for 16 hr in the second dimension (see below) gives a good separation of oligonucleotides which contain no uridylic acid residues but is less satisfactory for other graticules. Good resolution of oligonucleotides in the "one U" graticule was achieved by extending the duration of electrophoresis in the second dimension to 40 hours. Oligonucleotides containing two and three uridylic acid residues were best separated when RNA was digested with a combination of  $T_1$  RNase and bacterial alkaline phosphatase (" $T_1$  plus phosphatase") and then subjected to electrophoresis for 40 hours in the second dimension. Bacterial alkaline phosphatase dephosphorylates the guanidylate residue at the 3' terminus of oligonucleotides and increases their mobility during electrophoresis on DEAE paper (Fig. 2.2 c).

Samples were prepared for digestion by lyophilisation in siliconised test tubes. Digestion with  $T_1$  RNase was carried out with an enzyme to substrate ratio of one to ten or one to twenty in 0.02 M tris-HCl, 0.002 M EDTA, pH 7.4 for 30-45 mins. For  $T_1$  plus phosphatase digestion an enzyme to substrate ratio of one to twenty for both enzymes was used, digests being carried out in 0.02 M tris-HCl, pH 8.5 for 30-45 mins. Exact digestion conditions were varied with the activity of the particular enzyme preparations used.

Initial experiments were carried out on 40-50  $\mu$ g aliquots of RNA.

In later experiments 100-120  $\mu\text{g}$  amounts were digested. The greater amount of radioactivity applied facilitated subsequent detection and quantitation of material.

(b) Pancreatic Ribonuclease and  $T_1$  plus Pancreatic Ribonucleases.

Pancreatic ribonuclease cleaves RNA to yield oligonucleotides with either a uridylic acid or cytidylic acid at their 3'-terminus. Two dimensional separation of the products of this type of digestion gives fingerprints with two different sets of graticules; one set being oligonucleotides terminated by uridylic acid and the other set being oligonucleotides terminated by cytidylic acid. (Fig. 2.3).

RNA digestion with combined pancreatic and  $T_1$  RNases yields a number of short oligonucleotide products. The mobility of oligonucleotides in the second dimension of the fingerprinting system is decreased by increasing content of adenylic acid residues. Only a comparatively small number of spots are seen in this type of fingerprint. Many methylated spots are completely separated from non-methylated sequences and may be visualized in  $^{32}\text{P}$  fingerprints (Maden & Salim, 1974.).

Digestion with both pancreatic RNase and combined  $T_1$  plus pancreatic RNases is achieved by use of identical conditions to those already described for  $T_1$  RNase alone.

Digestion of amounts of RNA less than 50  $\mu\text{g}$  were carried out in a volume of 5  $\lambda$ . Amounts of RNA greater than this were digested in a volume of 10  $\lambda$ . All digestions were carried out in the tip of

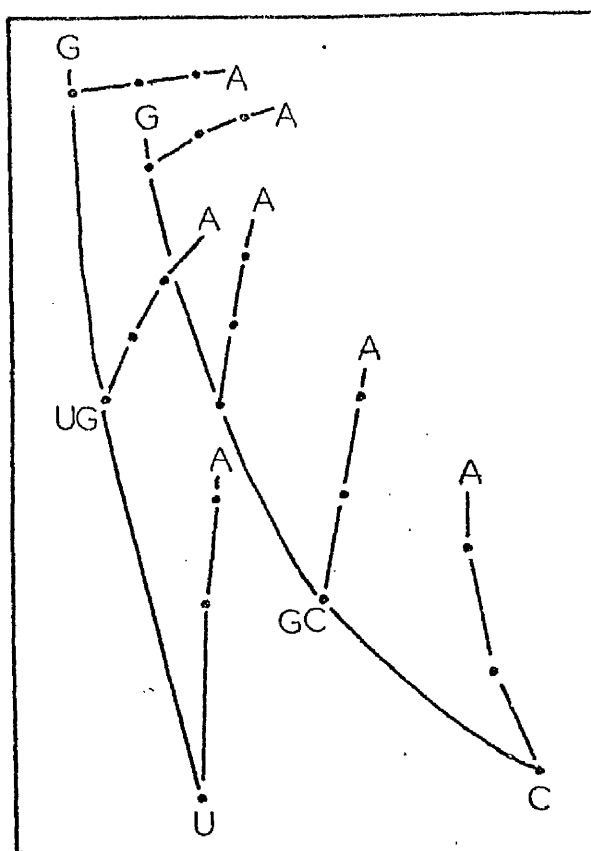


Fig.2.3 Fractionation of oligonucleotides after pancreatic  
RNase digestion.

a drawn-out capillary at 37°C in a humidified oven and were applied directly to the fractionating system.

#### 2.4.2 2-Dimensional Fractionation

The products of enzymic digestion of RNA were separated by two dimensional electrophoresis.

Separation in the first dimension was carried out on cellulose acetate at pH 3.5. For digests of less than 50 µg of RNA membrane form cellulose acetate was used. For larger amounts of material cellogel was used.

Cellulose acetate strips (2.5 x 95 cm) were wetted with 7M urea, pH 3.5 buffer (5% acetic acid, 7M urea, adjusted to pH 3.5 with pyridine) and excess buffer was dried off from an area about 10 cm from one end. The RNA digest was applied to this dried area in the centre of the strip. Marker dye was applied on either side of the sample spot. The whole strip was blotted and quickly placed over an electrophoresis rack in the electrophoresis tank with the end of the strip nearest to the sample dipping into the cathode compartment. Electrophoresis was carried out at 4.8 KV with the cellulose acetate strip drawing very little current.

For standard  $T_1$ , digests, pancreatic digests and  $T_1$  plus pancreatic digests the first dimension was generally run until the distance between the blue marker and slowest pink marker was about 35 cms. In the case of long  $T_1$  and  $T_1$  plus phosphatase

digests electrophoresis was continued until the slowest pink dye was 2 cm from the anode buffer. First dimension runs on cellogel strips were about 30% faster than runs on standard cellulose acetate strips.

On completion of the first dimension, electrophoresis strips were removed from the tank and excess white spirit was then allowed to drip off. The fractionated oligonucleotides were blotted onto a sheet of DEAE cellulose paper (DE81, 43 x 94 cm) at a distance of 10 cm from one end with a pad of 5 strips of Whatman 3 MM paper soaked in water. After allowing 30 mins. for transference of oligonucleotides the wetted area of the DEAE sheet was dipped in absolute alcohol and washed for about 2 mins. This process removes urea which is transferred from the cellulose acetate strip with the oligonucleotides.

For standard  $T_1$  digests the transfer was carried out with the blue marker dye on the cellulose acetate strip 4" in from the left-hand edge of the DEAE paper. Long  $T_1$  and  $T_1$  plus phosphatase oligonucleotides were blotted with the blue dye 1" from the left-hand edge. Pancreatic and  $T_1$  plus pancreatic separations were transferred with the original point of digest application about 1" from the left hand edge of the sheet.

For the second dimension of electrophoresis, marker dye was applied and one half of the paper carefully wetted with 7% formic acid. The sheet was draped over an electrophoresis rack and

the other half wetted. The rack was carefully lowered into the electrophoresis tank and electrophoresis carried out at 1,000-1,400 V. Electrophoresis was continued until the blue dye had travelled about 40% of the length of the paper for standard  $T_1$ , pancreatic and  $T_1$  plus pancreatic digests (16-20 hrs). For long  $T_1$  and  $T_1$  plus phosphatase separations electrophoresis was continued until the blue dye had travelled about 90% of the length of the paper (36-40 hrs).

After electrophoresis the paper was thoroughly dried in an oven until no smell of formic acid remained.

#### 2.4.3 Autoradiography.

The DEAE paper was marked with ink containing  $^{35}\text{S}$ -sulphate to correct orientation and was fixed to sheets of Kodirex X-ray film. Autoradiography was carried out in lead folders, kept in a light-tight cupboard. After exposure X-ray films were developed in DX-80 developer and fixed in FX-40 fixer. Generally 2  $\mu\text{Ci}$  of applied material gave satisfactory autoradiograms overnight.

#### 2.4.4 Recovery of Oligonucleotides from Fingerprints.

##### (a) Excision of Oligonucleotides

Autoradiographs were accurately aligned with the corresponding DEAE sheets using the markings made with  $^{35}\text{S}$  containing ink. The positions of the oligonucleotide "spots" were marked and the spots were excised from the sheet.

For quantitation of radioactivity in each spot the excised pieces of DEAE paper were placed in scintillation vials with 10 ml of PPO/toluene and subjected to liquid scintillation counting. In the case of spots covering a large area the DEAE paper was cut into several pieces and each piece counted separately.

#### (b) Elution of Oligonucleotides

Oligonucleotides are bound ionically to DEAE paper and it is necessary to neutralize the charge on the DEAE in order to elute them. This was done by eluting with a 30% aqueous solution of triethylaminecarbonate, pH 10. Elution from Whatman 52 paper was done with water as there are no ionic charges to neutralize in this case. Elution was continued until 0.2 - 0.4 ml had been collected on a polyvinyl chloride sheet. Eluates were evaporated to dryness in an oven at 60°C.

Triethylamine carbonate was made by adding pieces of solid CO<sub>2</sub> to a 30% solution of aqueous redistilled triethylamine until the solution became clear. The pH was checked to be within the range pH 9.9 - 10.1

### 2.5 Analysis of oligonucleotides and RNA

#### 2.5.1 Complete Hydrolysis of Oligonucleotides and RNA

For base composition analysis and for separation of pseudouridine from uridine, oligonucleotides and whole RNA species were completely hydrolysed. In early experiments hydrolysis was carried out using digestion with alkali whereas



in later experiments digestion was routinely carried out using  $T_2$  ribonuclease in conjunction with  $T_1$  and pancreatic ribonucleases.

Alkaline Hydrolysis (Brownlee 1972 pp 76-77)

Oligonucleotides were taken up in 10  $\mu$ l of 0.2 NaOH and sealed into drawn out capillary tubes. Hydrolysis was effected by incubating at 37°C for 16-18 hr. The hydrolysate was applied to Whatman No. 52 paper as a 1 cm streak and then subjected to electrophoresis at pH 3.5 for 40 mins. at 4.8 KV. Mononucleotides were detected by autoradiography and quantitated by liquid scintillation counting.

Hydrolysis with  $T_2$  RNase (Brownlee, 1972 pp. 210-211)

The substrate was taken up in 10  $\mu$ l of a solution containing 2.0 units per ml  $T_2$  - RNase in 0.05M ammonium acetate, pH 4.5 containing 0.05 mg each of  $T_1$  - and Pancreatic RNase per ml. Digestion was carried out in unsealed capillaries for 2 hrs. at 37°C in a humidified oven. Digestion products were then separated, detected and quantitated as described for alkaline hydrolysates.

Hydrolysis with  $T_2$ -RNase yields only 3'- phosphates whereas alkaline hydrolysis results in a mixture of 2'- and 3'- phosphates. Because of this, electrophoretic separation is much sharper after  $T_2$  -RNase hydrolysis (Fig. 2.4a). When only base composition data was required this gave no detectable alteration in quantitations. However, when studying the

presence of minor components such as pseudouridine it was found that  $T_2$  digestion gave more reproducible results (see results section). However, in this method it was essential to use only a few micrograms of substrate in 10  $\mu$ l of digestion mix, to ensure complete digestion.

## 2.5.2 Detection and Quantitation of Pseudouridine

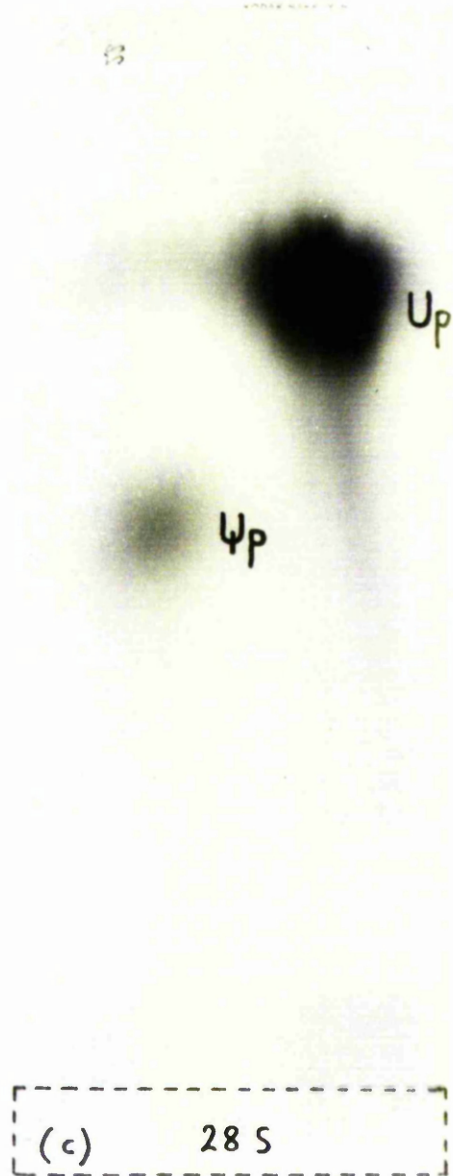
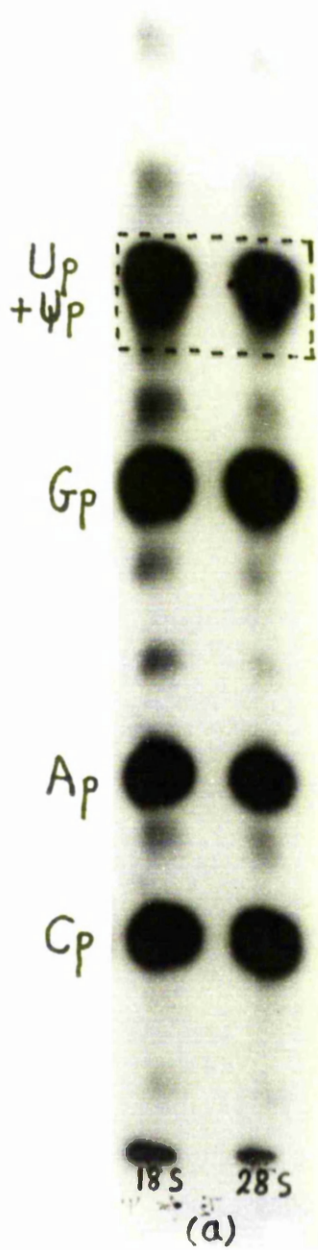
### (a) Chromatography, first dimension

$\psi_p$  is not significantly resolved from Up by electrophoresis at pH 3.5. Therefore the area containing Up and  $\psi_p$  was excised and stitched to a second sheet of Whatman 52 paper 15 cm from one end. Descending chromatography was carried out for 40 hours in isopropanol: HCl:  $H_2O$  (68:17.6: 14.4 by volume; Wyatt, 1951; Fellner, 1969). Fig. 2.4b shows an autoradiograph of such a separation. For each sample there is a heavily darkened area corresponding to Up and a less dark area corresponding to  $\psi_p$ . Between these is a diffuse streak due to material trailing behind the main Up spot. This trailing was variable in amount but could not be eliminated. Control experiments with pure Up indicated that for heavily labelled samples at least 2% of Up trailed into the  $\psi_p$  area. This degree of trailing was considered acceptable in the analysis of single oligonucleotides for the presence or absence of  $\psi_p$  when only approximate quantitation was required. However, when attempting accurate quantitative analysis of whole RNA samples, in which  $\psi_p$  is present to the extent of only a few percent of the amount of Up an additional chromatographic stage was carried out.

Fig. 2.4

Consecutive stages in quantification of  $\Upsilon$ p.

(a)  $T_2$  ribonuclease hydrolysates of 18 S and 28 S RNA were separated by electrophoresis on Whatman 52 paper in 5% acetic acid: pyridine buffer, pH. 3.5, for 40 min at 4.8 KV. The positions of the four mononucleotides are indicated. The faint products are alkali-stable compounds; see legend to Table 1 and Maden & Salim (1974). A very small amount of material has remained bound to denatured enzyme at the origin; this did not occur after alkaline hydrolysis. (b) The region which is indicated by the interrupted line in (a) was excised, stitched to another sheet of Whatman 52 paper and chromatographed as described in Methods.  $\Upsilon$ p displays an " $R_U$ " of 0.8 (Fellner, 1969). There is a trace of  $G_m - G_p$  near the origin. This sometimes overlaps  $\Upsilon$ p on electrophoresis but migrates slowly on chromatography. (c) The region indicated by an interrupted line in (b) for 28 S RNA, and a similar region for 18 S RNA (not shown), were excised, stitched to another sheet of chromatography paper and chromatographed at right angles to the first dimension. The second separation (vertical axis) was slightly longer than the first (left to right).



(b) Chromatography, second dimension

To complete the separation of  $\psi$ p and Up, the area shown in Fig. 2.4.b was excised and stitched to another sheet of Whatman 52 paper. Descending chromatography was repeated for 40 hours at right angles to the first dimension, using the same solvent. Fig. 2.4 C shows the result. Up now appears as a heavily darkened spot with arms trailing in two directions at right-angles to one another. The material which trailed in the first dimension has run normally in the second, indicating that this is Up and not a different compound or a breakdown product.  $\psi$ p, having migrated more slowly in both dimensions, is completely separated from Up and its trailing arms. For quantitation the regions of the chromatogram corresponding to Up and  $\psi$ p were first divided by visual inspection. Each region was then cut into sufficient 2 cm x 2 cm squares for scintillation counting of all labelled material, including the trailing arms of Up. To check the spillover of Up into the  $\psi$ p area a control experiment was carried out. The Up plus  $\psi$ p area from electrophoretic separation of a  $T_2$  digest was subjected to chromatography in one dimension. The main Up spot was cut out, excluding both trailing material and  $\psi$ p, and was subjected to chromatography in two dimensions as already described. When the chromatogram was quantitated it was found that the amount of Up in the supposed  $\psi$ p area was very small (see results).

Spectroscopic Measurements on Uridine and Pseudouridine

Cytoplasmic RNA was extracted from approximately

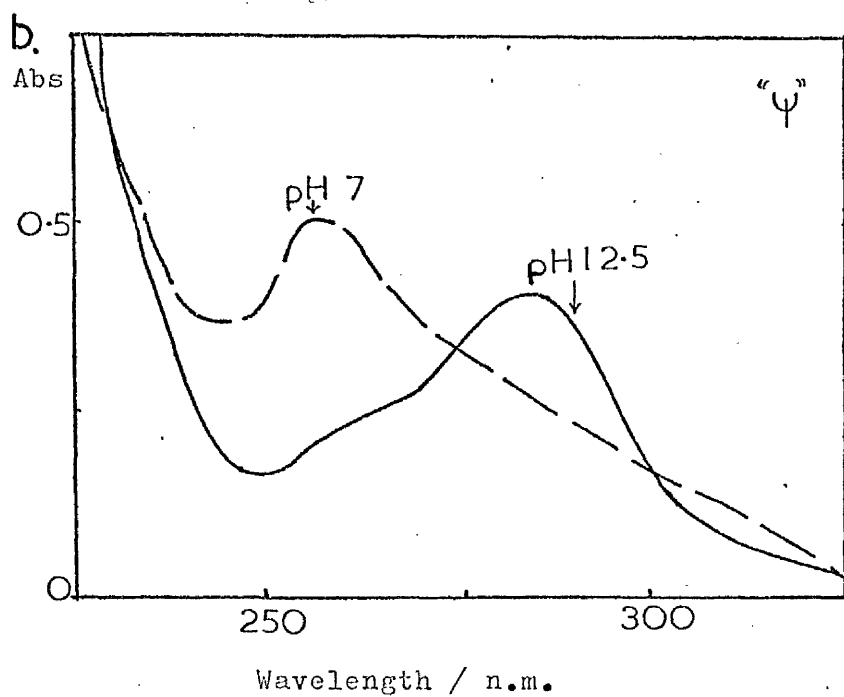
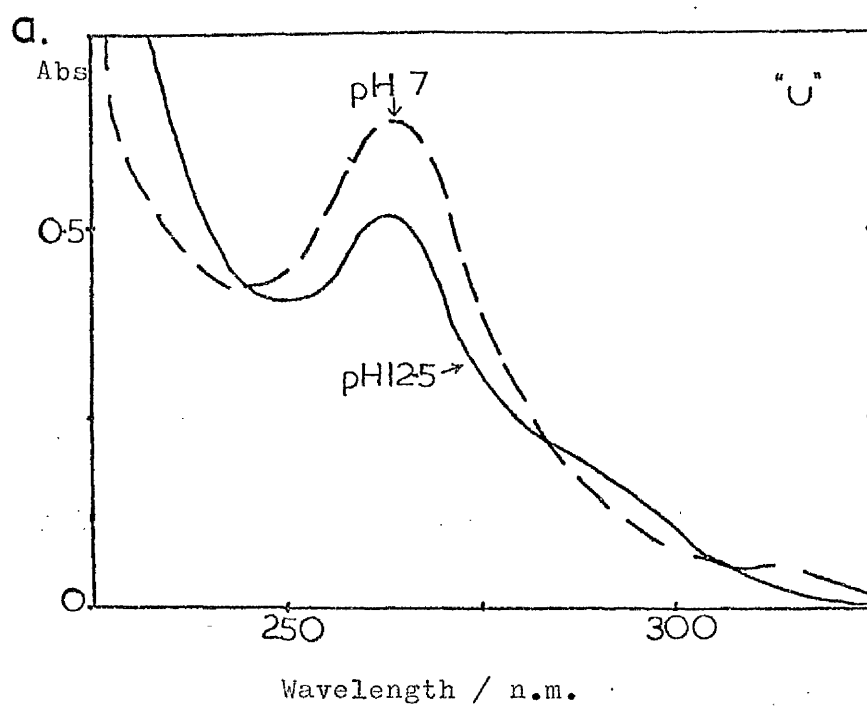


Fig. 2.5 Absorption spectra of (a)"uridine"  
and (b) "pseudouridine".

$5 \times 10^8$  unlabelled HeLa cells. This unfractionated RNA was digested with 0.5 ml of standard  $T_2$  digestion mix for 2 hours at  $37^\circ\text{C}$  and then applied as a 5 cm streak on Whatman 3 mm paper. Electrophoresis was carried out at pH 3.5 for 30 mins. at 4.8 KV. After drying the paper was observed under ultraviolet light to visualize the position of mononucleotides. The fastest moving band of material was cut out and designated "uridine". A fainter band moving just behind uridine was cut out and designated "pseudouridine". Material from both bands was eluted with triethylamine carbonate and freeze dried. The dried samples were taken up in 1 ml each of 0.01 M phosphate buffer pH 7.0 and their ultraviolet spectra scanned on an SP8000 spectrophotometer. 100  $\mu\text{l}$  of 10 N NaOH was added to each sample and their u. v. spectra were again scanned. (see Fig. 2.5). Addition of 0.1 volumes of 10 N NaOH to 0.01 M phosphate buffer was found to alter its pH from 7.0 to 12.5

### 2.5.3 Digestion of Oligonucleotides with Pancreatic RNase

It was useful to digest  $T_1$  and  $T_1$  plus phosphatase oligonucleotides containing adenylate residues with pancreatic RNase.

Samples were digested in 10  $\mu\text{l}$  of 0.01 M tris-HCl, 0.001 M EDTA, pH 7.4 containing 0.2 mg/ml pancreatic RNase. Digestions were carried out for 30 mins. at  $37^\circ\text{C}$ . Separation of products was achieved by electrophoresis on Whatman 52 paper of pH 3.5. Products were identified either by their mobility (Brownlee, 1972)

or by base composition analysis.

#### 2, 5.4. Complete Digestion with Snake Venom Phosphodiesterase

This type of digestion was useful in determining the 5'-terminal nucleotide of oligonucleotides of known base composition. Digestion was carried out in 10  $\mu$ l of 0.025 M tris-HCl, 0.01 M Mg acetate, pH 8.5 containing 0.1 mg/ml of enzyme. Incubation was for 2 hrs. at 37°C. The mononucleotides were separated by electrophoresis at pH 3.5 on Whatman 52 paper.

#### 2.6 Preparation and Properties of Fragments of HeLa 28S rRNA

##### 2.6.1 Partial Digestion of 28S rRNA with $T_1$ RNase

28S rRNA was digested with  $T_1$  RNase under mild conditions based on those described by Cox et al (1973). 100  $\mu$ g samples of rRNA were incubated for 1 hour at 20°C in 80  $\mu$ l 0.01M phosphate buffer pH 7.0 containing 0.1 units of  $T_1$  RNase. The reaction was stopped by adding 0.7 ml LETS and cooling on ice. Samples were layered onto 10 - 25% sucrose LETS gradients in Beckman SW 40 buckets and were centrifuged at 39,000 r. p. m. for 18 hours at 20°C. Gradients were harvested and optical density at 260 m. m. monitored by pumping through the Gilford spectrophotometer. The flow rate was set such that about 30 fractions were collected per gradient (see fig. 2.6). Gradient fractions marked A were pooled and 20 - 30  $\mu$ g cold carrier rRNA was added. RNA was precipitated with  $2\frac{1}{2}$  volumes of ethanol.



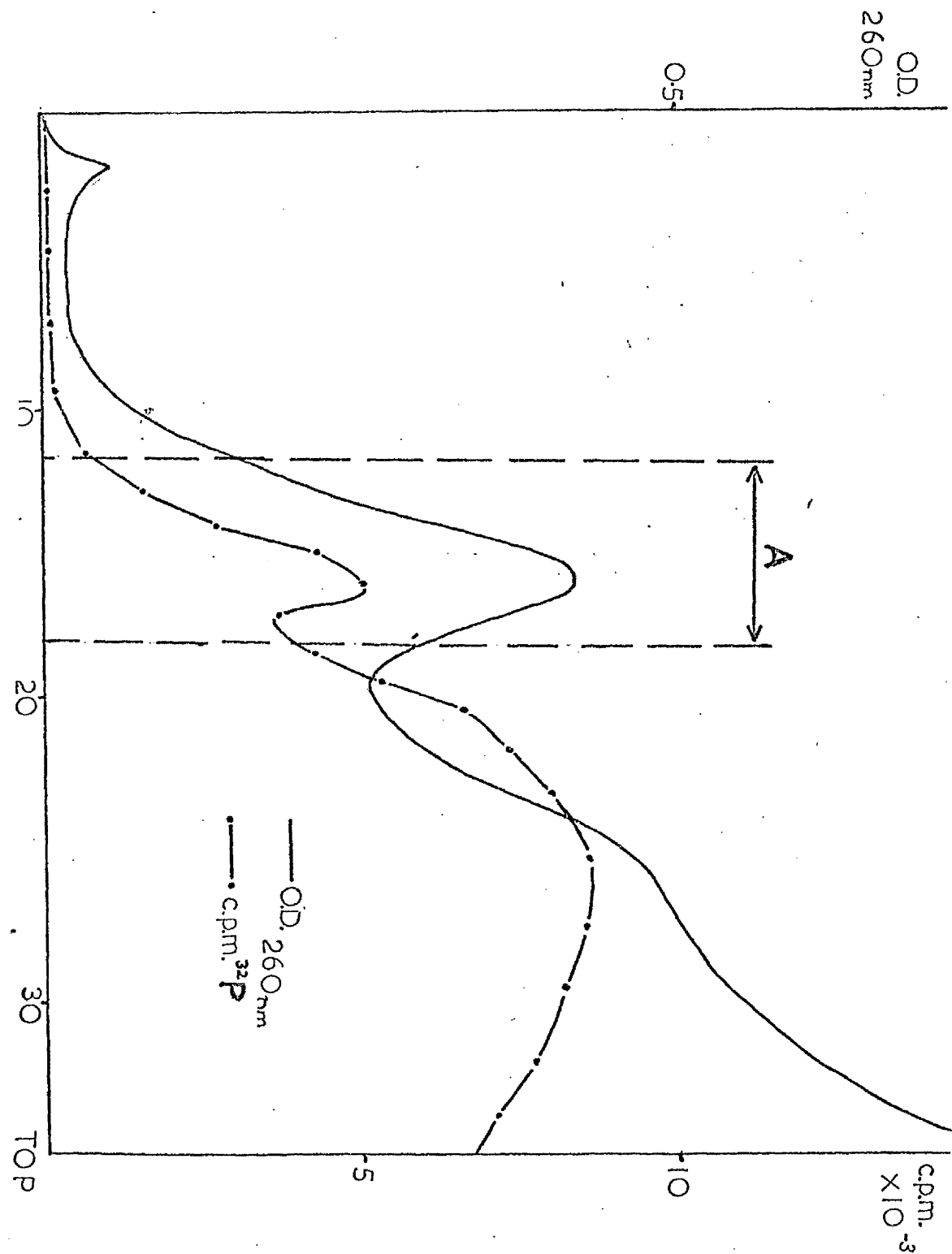


Fig 2.6 O.D. profile of products of partial digestion of HeLa 28S rRNA after centrifugation on 10-25% sucrose-LETS gradients (conditions described in text).

After one ethanol precipitation fragment A was purified by rerunning on 15 - 30% LETS sucrose gradients for 18 hrs. as previously described. A single peak of absorption at 260 n.m. was observed when gradients were fractionated. The fractions directly under the O.D. peak were pooled and carrier RNA was added. Preparations were precipitated twice with ethanol.

$^{32}\text{P}$  labelled 28S RNA was partially digested with  $\text{T}_1$  RNase and the products separated on 15 - 30% sucrose gradients as described above. 10  $\lambda$  aliquots from each fraction were taken and assayed for radioactivity in 10 ml toluene/triton scintillation fluid in a Packard Tricarb scintillation counter. Fig. 2.6 shows the distribution of radioactivity across the gradient.

#### 2.6.2 Polyacrylamide Gel Electrophoresis

Fragment A preparations were analysed by electrophoresis on polyacrylamide slab gels (30 x 15 cm) as described by De Wachter & Fiers (1971).

Slab gels were poured with the following constituents:-  
6% acrylamide, 6, 0.2% NN';Methylenebisacrylamide, 0.015% SDS,  $1.5 \times 10^{-6}$  M EDTA, 0.08% Ammonium Persulphate, 0.04%, 0.04% TEMED, 40 mM tris-acetic acid pH 8.0. RNA samples were dissolved in 0.05 ml of water and an equal volume of dye/sucrose mixture was added. The samples were loaded into the slots at the top of the gel with a syringe.

Electrophoresis was carried out with a 40 mM tris-acetic

acid buffer system, pH 8.0 at 25-30 mA for 20-24 hours at 4°C.

Gels were autoradiographed using X-ray film in a similar manner to that described for electrophoretic separations on paper.

The main bands, visualized by autoradiography, were cut from the gel for further investigation on denaturing gels containing 6M urea. In this case gels contained 9% acrylamide, 0.3% NN'-methylenebisacrylamide, 0.025% SDS,  $2.25 \times 10^{-6}$  M EDTA, 6M urea, 25 mM citric acid, 0.01 g/l ferrous sulphate, 0.15 g/l hydrogen peroxide and 0.04% ascorbic acid.

Pieces from the first-dimension gel were set at the top of the denaturing slab in a stacker gel of the same composition as the slab except for the absence of citric acid. Dye/sucrose mixture was applied on top of this stacker and electrophoresis was carried out using 6M urea buffer at 15 - 20 mA for 16-18 hrs. at 4°C. Autoradiography was then carried out as previously described.

### 2. 6. 3 Electron Microscopy of Fragment A

Samples of fragment A were spread for electron microscopy with the direction and assistance of Dr. Lesley Coggins of the Beatson Institute for Cancer Research. A modified Kleinschmidt technique was used based on those described by Robberson et al., (1971), and by Wellauer and Dawid(1973).

Carbon films were prepared by stripping from freshly cleaved mica sheets onto distilled water. Films were lowered onto 400-mesh copper grids which had been dipped in

1% polybutene in xylene to improve adhesion.

Fragment A was dissolved to a concentration of 1  $\mu\text{g/ml}$  in 200  $\mu\text{l}$  of spreading solution (hyperphase) containing 4 M urea, 80% formamide, 25 mM EDTA (pH 8.5) and 30  $\mu\text{g/ml}$  cytochrome C (Sigma, type VI) and allowed to stand for 10 mins.

The hyperphase was spread on the surface of the distilled water hypophase by running the solution slowly down a glass rod. The surface film was detected by dusting the hypophase with talcum powder and noting the appearance of a clear area when the hyperphase was spread. Carbon coated grids were dipped in ethanol to make them hydrophilic and then touched onto the surface of the film. Each grid was dried on filter paper and stained by placing in  $5 \times 10^{-5}$  M uranyl acetate in 90% ethanol (Davis et al., 1971) for 5 mins. Samples were then dehydrated in absolute ethanol and rotary shadowed with Pt-Pd (80:20) at an angle of  $10^\circ$ .

Grids were scanned under the electron microscope and a number of fields were photographed.

#### 2.6.4. Bisulphite Modification of RNA

RNA was precipitated from ethanol by centrifugation and the pellet drained. The sample was dissolved in 10 mM  $\text{MgCl}_2$  3M sodium metabisulphite, pH 6.0. An aliquot was withdrawn immediately and diluted ten times with 10 mM  $\text{MgCl}_2$  on ice to stop reaction. The remainder of the sample was incubated at

25°C with further aliquots being withdrawn at suitable intervals.

After the final aliquot had been withdrawn the removal of bisulphite and destruction of bisulphite adducts was carried out as follows:-

(i) Bisulphite was removed by dialysis against 0.15 M NaCl, 20 mM tris-HCl, pH 7.5 and then against 0.15M NaCl, 5 mM tris-HCl, pH 7.5 for two hours each at 4°C.

(ii) Removal of bisulphite adducts was effected by dialysis against 0.1M Tris-HCl- pH 9.0 at 37°C for 18 hours.

(iii) RNA was neutralized by dialysis against 20 mM tris-HCl pH 7.0 then 5 mM Tris-HCl pH 7.0 for 2 hours each at 4°C.

(iv) Finally the samples were dialysed extensively against water at 4°C to remove salt.

### Chapter 3

#### The Pseudouridylic Acid Content of Eukaryotic rRNA.

Pseudouridine has attracted little general attention as a modified nucleoside in rRNA. This may be due, at least in part, to analytical difficulties such as the lack of a specific radioactive label for pseudouridine. The pseudouridine content of RNA's is usually examined by complete alkaline or enzymic hydrolysis to mononucleotides (and alkali-stable, 2'0 methylated, dinucleotides), followed by chromatographic separation of the products. Methods of separating pseudouridine 3' (2') phosphate ( $\Psi$ p) from uridine 3' (2') phosphate (Up) include chromatography on ion exchange columns (Glitz and Decker, 1963; Amaldi & Attardi, 1968) or on Whatman paper (using the system of Wyatt, 1951 as reported by Fellner, 1969, or the system of Singh and Lane, 1964) or on thin layers (Nishimura, 1972; Rubin, 1973). The procedure of Glitz & Decker (1963) yielded excellent resolution of  $\Psi$ p from Up, but the large number of fractions which were obtained rendered the method laborious. The method used here was adapted from the micromethods already in use for RNA sequence analysis (Brownlee, 1972). It consisted of alkaline or  $T_2$  ribonuclease digestion of the RNA, base composition analysis by electrophoresis at pH 3.5, and two dimensional paper chromatographic separation of  $\Psi$ p from Up using the solvent system of Wyatt (1951).

### 3.1 HeLa Cell rRNA

#### 3.1.1, Base Composition

Table 3.1 shows the base composition data for HeLa cell rRNA. The values are means of several alkaline and  $T_2$  hydrolysates (Table 3.1 a). Agreement between the results obtained by the two methods is close and there is also good agreement with published values (Table 3.1c). From estimates of the chain lengths of the RNA molecules and allowing for small quantities of material released as alkali-stable products (for both points see notes to Table 3.1), it is possible to estimate the molar amounts of Up plus  $\psi$ p released as mononucleotides in hydrolysates of the RNA's (last column of table 3.1. b).

#### 3.1.2 Pseudouridine content

##### One Dimensional Chromatography

Quantitation of  $\psi$ p in HeLa rRNA was first attempted by separating  $\psi$ p and Up from the Up plus  $\psi$ p electrophoretic band by chromatography in one dimension using the solvent system of Wyatt (1951). Up trails to a small but variable extent in this system. Consequently the standard deviation of the values obtained was quite large and the blank values for previously purified Up were very variable (Fig. 2.4). This made accurate quantitation of  $\psi$ p impossible. Two-dimensional chromatography as described under methods, overcame the problem of trailing and was used for more accurate estimation of the pseudouridine

Table 3.1

Nucleotide Compositions of HeLa rRNA's

(a) Comparison of Alkali and T<sub>2</sub> Digests

<u>18S</u>	%				Number of determinations
	Cp	Ap	Gp	Up + Up	
Alkali	26.5	22.9	28.7	21.9	4
T <sub>2</sub>	26.6	23.5	28.7	21.2	6

<u>28S</u>	%				Number of determinations
	Cp	Ap	Gp	Up + Up	
Alkali	33.1	16.4	35.3	15.2	3
T <sub>2</sub>	32.3	17.2	34.3	16.2	6

(b) Combined Alkali and T<sub>2</sub> Results

	Cp	Ap	Gp	Up + Up	Total mono- nucleotides (i)	Molar Amount Up + Up
18S	26.5	23.2	28.7	21.5	1950	420
28S	32.6	16.9	34.6	15.9	5000	795
5.8S	30.9	18.7	28.3	22.1	155	34.3

5.8S results are from 5 determinations using T<sub>2</sub> ribonuclease.  
 18S and 28S results are a combination of results from determinations  
 using T<sub>2</sub> ribonuclease or alkaline hydrolysis. 28S determinations  
 were for 5.8S -free 28S RNA.

(c) Results from Other Sources

	Cp	Ap	%	Gp	Up + Up	Source
18S	27.1	23.0		28.6	21.3	(ii)
	27.3	20.2		30.5	22.0	(iii)
	Cp	Ap	%	Gp	Up + Up	Source
28S	32.0	16.5		35.2	16.3	(ii)
	32.3	16.0		35.0	16.7	(iii)
5.8S	26.3	21.5		29.8	22.4	(iv)



### Notes on Table 3.1

(i) The total numbers of mononucleotides per RNA were estimated from published molecular weight values and sequence data as follows. 18S RNA:- molecular weight from hydrodynamic data (Peterman & Pavlovec, 1966),  $0.67 \times 10^6$ , from electron microscopic contour lengths (Wellauer and Dawid, 1973),  $0.67 \times 10^6$  (with reference to E. coli rRNA, whose molecular weight was determined by hydrodynamically by Stanley and Bock, 1965); mean 18S value,  $0.67 \times 10^6 = 2010$  nucleotides.

Of these approximately 76 would be released as alkali-stable dinucleotides, of which 60 would migrate differently from the mononucleotides on electrophoresis and the remaining few would overlap, and be included with, Ap or Gp (Maden & Sálím, 1974). Therefore the label scored as mononucleotides in the hydrolysates = 1950 nucleotides, of which Up +  $\bar{U}$ p = 21.58, or 420. 5.8S free - 28S RNA:- "hydrodynamic" molecular weight (Petermann and Pavlovec, 1966)  $1.64 \times 10^6$ ; electron microscopy (Wellauer and Dawid, 1973)  $1.76 \times 10^6$ ; mean,  $1.70 \times 10^6 = 5,100$  nucleotides. Subtract 76 for resolved, alkali-stable material:- approximately 5,000 migrating as mononucleotides. 5.8S:- The rat hepatoma sequence contains 158 nucleotides (Nazar et al., 1975). The HeLa sequence is probably identical except for an extra U at the 3' end (Maden & Robertson, 1974). Of the 159 nucleotides one is released as  $U_{OH}$ , two are released as  $G_{mp}C_p$  and fractional amounts as  $U_{mp}G_p$

and  $C_p$ . Therefore 155 are released as mononucleotides.

The base composition data, which indicate  $34.3 U_p + \psi_p$  residues

are in good agreement with the sequence data, which indicates

$34.8 U_p + \psi_p$  residue. (0.2 mole  $U_p$  is released as  $U_{mp} G_p$

(Khan and Maden, 1976).

(ii) From Willems et al (1968).

(iii) From Amaldi & Attardi (1968)

(iv) From Pene et al (1968).

content of rRNA.

### Two Dimensional Chromatography.

Table 3.2 shows the data for pseudouridine analysis after two-dimensional chromatography on material from Up plus Up electrophoretic bands (Fig. 2.4). Only data for  $T_2$  hydrolysates are shown. These were highly reproducible, whereas the alkaline hydrolysis results were less so. This is probably because of the presence of 2' and 3' phosphates, causing somewhat less sharp separations than in the  $T_2$  hydrolysates. The molar amounts of Up were then calculated from the above values and those in the last column of table 3.1b. The results are shown in column 2 of table 3.2. Values from previously published analyses are shown in parentheses for comparison (column 3).

Three pseudouridines are not liberated as free Up in the digests. One of these occurs in a hypermodified nucleotide in 18S RNA (Saponara & Enger, 1974; Maden et al., 1975) and the other two are found in alkali stable compounds (Um-G and Um-Gm-U) within 28S RNA (Maden & Salim, 1974). Column 5 gives the final estimates of pseudouridine in HeLa cell rRNA corrected to include these extra pseudouridines.

The standard deviation for 5.8S RNA is larger than that for the high molecular weight rRNA's. However, this represents only a small variation in the estimations of the molar amount of Up. Although greater molar amounts of 5.8S RNA were

TABLE 3.2.

	$\frac{\psi p}{Up + \psi p} \%$	Molar amount $\psi p$	(previous data)	plus (d)	total $\psi$	Number of determinations
18 S	$8.59 \pm 0.14$	$36 \pm 1$	(~29) (b)	1	$37 \pm 1$	5
28 S	$7.29 \pm 0.23$	$58 \pm 2$	(~55) (b)	2	$60 \pm 2$	5
5.8 S	$7.75 \pm 0.85$	$2.7 \pm 0.2$	(<2) (c)	-	$2.7 \pm 0.2$	5
Blank (a)	0.13	-	-	-	-	3

Pseudouridylate content of HeLa cell rRNA, determined as described in the text.

(a) The blank value was obtained by rerunning previously purified Up. This value has been subtracted from the other values. (b) The values in parenthesis for 18 S and 28 S are calculated from data of Amaldi and Attardi (1968), who recorded a base composition of 1.5%  $\psi p$  for 18 S RNA and 1.1%  $\psi p$  for 28 S RNA, using column chromatographic separations. (c) Two  $\psi p$  residues were previously identified in the 5.8 S sequence, one of them in slightly submolar yield (Nazar et al 1975; Maden & Robertson, 1974), see text for further discussion. (d) The nature of these extra pseudouridylates is indicated in the text.

used the total radioactivity applied to chromatograms was much less than for 18S and 28S RNA's. Liquid scintillation counting of  $\psi$ p in 5.8S rRNA involves quantitation of a small percentage of a fairly small number and is not as accurate as when larger amounts of radioactivity are being measured.

Identification of the minor component of the chromatograms as pseudouridylic acid is based on three main lines of evidence.

(1) The "Ru" value (0.8) is approximately that reported for  $\psi$ p in this chromatographic system (Fellner, 1969). Other modified uridylates migrate differently: methylated uridylates more rapidly than Up, and dihydrouridylic acid (Ru=0.9) is unstable in alkali (Brownlee, 1972, p.205).

(2) Spectra of "Up" and of " $\psi$ p" were measured at pH 7 and pH 12.5 (see fig. 2.3). The spectrum of "uridine" is unchanged whilst that of "pseudouridine" shows a shift of  $\lambda$  MAX from 262 nm to 286 nm as is expected for pseudouridine (Cohn, 1960). The spectrum of  $\psi$ p shows a slight shoulder of absorption at 262 nm suggesting that some contaminating Up may be present. This is consistent with results of one-dimensional chromatographic separations of radioactive material where trailing of Up into  $\psi$ p was observed.

(3) Larger scale analysis of  $\psi$ p in eukaryotic rRNA (Amaldi & Attardi, 1968; Glitz & Decker, 1963) gave roughly comparable molar yields.

The two-dimensional separation of  $\psi_p$  from  $U_p$  seems to be essentially complete based on the following criteria:-

(i) the results are highly reproducible as shown by the small standard deviation, particularly for high molecular weight rRNA.

(ii) the "pure U" control gives very little spillover into the area of the chromatogram designated as containing  $\psi_p$ .

The main advantages of the method of  $\psi_p$  determination used are:-

(1) The method is suitable for small quantities of radioactively labelled material. (ii) At no stage is material eluted from the Whatman paper. Therefore there is no possibility of selective losses during manipulations. (iii) Because  $U_p$  runs fairly rapidly relative to the solvent front ( $R_f = 0.8$ ), the system favours maximal separation of  $U_p$  from  $\psi_p$ . Therefore, by repeating the separation in a second dimension, using the same solvent system, it is possible to obtain complete separation of relatively small quantities of  $\psi_p$  from much larger quantities of  $U_p$ .

#### Other Solvent Systems

A further solvent system commonly used for separating minor components of RNA is composed of propan-2-ol(70), water (30) and ammonia (1) (Markham & Smith 1952, as described by Brownlee, 1972). In this system  $\psi_p$  has an  $R_u$  of 0.63 ( $R_u$  of  $\psi_p$  in propan-2-ol/HCl/water is 0.79). However, most nucleotides have very low  $R_f$ 's in this system.

Two dimensional separation of  $\psi$ p from Up was attempted using propan-2-ol/HCl/Water in the first dimension and propan-2-ol/ $\text{NH}_3$ /water in the second. Up and  $\psi$ p both move a considerable distance in the first dimension (Up has Rf of 0.8). The low Rf's of both species in the second dimension means that the separation of Up and  $\psi$ p is not greatly improved by the use of this second dimension. Therefore this system was not used further.

### 3.1.3 Pseudouridine and 2'-O-methylation

From the present pseudouridine determinations and from previous quantitation of the methylated sequences in HeLa cell rRNA (Maden & Salim, 1974), Lane's (1965) suggested correlation between pseudouridine content and numbers of 2'-O-methyl groups can be re-examined. The data are shown in table 3.3. The overall correlation is remarkably close in the case of the major cytoplasmic rRNA's.

### 3.2 The Pseudouridine Content of rRNA from Other Eukaryotes

The numerical correlation between pseudouridine content and 2'-O-methyl groups in HeLa rRNA is sufficiently close to warrant attention. Investigations were carried out to determine whether this correlation is also true for the rRNA's of other eukaryotes. The pseudouridine content of rRNA of three other species was determined. The species chosen were, a further mammal (mouse L-cells), a lower vertebrate (*Xenopus laevis*) and a non-vertebrate eukaryote (*Dictyostelium discoideum*).

TABLE 3.3

Comparison of the Numbers of Pseudouridylate Residues and  
2' - 0 - methyl Groups in HeLa - Cell r RNA.

	Total $\psi$	Total 2'-0-methyl groups
18 S	37	38
28 S	60	62
5.8 S	2.3	1.2

The methylation data are based on Maden & Salim (1974), with very minor refinements from a re-analysis by the " $T_1$  plus pancreatic" ribonuclease fingerprinting system (Maden & Khan, 1976), and the pancreatic ribonuclease system (B. E. H. Maden, unpublished results). 5.8 S RNA contains one unimolar and one fractional (0.2 molar) 2'-0-methyl group.



### 3.2.1 Base Composition

Tables 3.4, 3.5 and 3.6 show the base composition data for eukaryotic 18S, 28S, and 5.8S RNA's. The values are means of several  $T_2$  hydrolysates. Results obtained for high molecular weight rRNA are in reasonable agreement with those published for L-cells (Lane & Tamaoki, 1969), *Xenopus laevis* (Birnstiel et al., 1968) and *Dictyostelium discoidium* (Jacobson et al., 1974). No published data are available on the exact base-composition of 5.8S rRNA from L-cells, *Xenopus* or *Dictyostelium*. However, the base compositions of all the 5.8S rRNA's studied appear to be related in the same way to the composition of the 28S RNA's from which they are derived. Approximate molar amounts of Up plus  $\Psi$ p released as mononucleotides were calculated from estimated chain lengths of the RNA molecules (See table legends).

The exact number of methyl groups in *Xenopus* rRNA's has recently been determined (B. E. H. Maden, unpublished observations). *Xenopus* 18S RNA contains some 33 ribose- and 7 base- methylations whilst 28S RNA contains 66 ribose- and 5 base-methylations. These figures were used when calculating numbers of mononucleotides released from *Xenopus* RNA's.

The study of *dictyostelium* rRNA was only commenced recently. Data are not available either on exact molecular weights or the degree of methylation of rRNA in this species. Extensive data

TABLE 3.4.

<u>Nucleotide Compositions of 18S rRNA.</u>					Total mono nucleotides	Molar amount Up + $\Psi$ p
<u>Species</u>	<u>%</u>					
	Cp	Ap	Gp	Up + $\Psi$ p		
HeLa (a)	26.5	23.2	28.7	21.5	1950	420
L-cell (b)	26.7 (27.5)	23.5 23.0	29.9 29.8	19.9 19.3)	1960	390
Xenopus (c)	25.4 (24.3	21.4 23.5	29.9 28.8	23.3 23.4 )	1960	450
Dictyostelium (d)	16.9 (18.3	30.4 28.9	24.3 23.6	28.4 29.2)	2040	580

(a) Determination of HeLa base composition was outlined in the legend to table 3.1. Base compositions of 18S RNA of other species are means of 4 determinations using T<sub>2</sub> ribonuclease.

(b) Figures in parenthesis from Lane & Tamaoki (1969). Molecular weight of L-cell 18S RNA is taken as being the same as HeLa on the basis of electron microscopic measurements (Wellauer & Dawid, 1973; Wellauer et al, 1974) 2010 nucleotides. Of these approximately 70 are released in alkali-stable dinucleotides of which 50 would migrate separately from mononucleotides (Hashimoto et al 1975). Label scored as mononucleotides = 1960.

(c) Figures in parentheses from Birnstiel et al (1968). Molecular weight taken from electron microscopic measurements is similar to HeLa (Wellauer & Dawid, 1974). The number of methyl groups is slightly less than in HeLa 18S RNA (B. E. H. Maden, unpublished observation). Approximately 50 nucleotides would migrate separately from mononucleotides.

(d) Figures in parentheses from Jacobson et al (1974). Precise data are not available on the molecular weight and degree of methylation of Dictyostelium rRNA. Figures are available for yeast which is a closely related species and these are assumed to be approximately correct for Dictyostelium (Klootwijk & Planta, 1974; Loening, 1968). Of a total of about 2065 nucleotides 36 occur in alkali stable dinucleotides. Approximately 32 run separately from mononucleotides.

TABLE 3.5.  
Nucleotide Compositions of 28S rRNA

<u>Species</u>	Cp	% Ap	Gp	Up + Up	Total mono- nucleotides	Molar amount Up + Up
HeLa (a)	32.6	16.9	34.6	15.9	5000	795
L-Cell (b)	30.5 ( 30.8	18.4 17.5	36.6 35.3	14.5 16.0 )	5000	725
Xenopus (c)	28.6 ( 28.3	19.3 19.2	34.8 35.4	17.3 17.1 )	4600	795
Dictyostelium (d)	17.6 (17.6	28.5 28.0	25.6 25.3	28.5 29.1 )	3750	1070

(a) Determination of HeLa base composition was outlined in the legend to table 3.1. Base compositions of 28S RNA of other species are means of 4 determinations using T<sub>2</sub> ribonuclease.

(b) Figures in parentheses from Lane and Tamaoki (1969). Molecular weight for L-cell 28S is taken as being the same as HeLa on the basis of electron microscopic measurements. Approximately 80 nucleotides run in resolved alkali-stable species. (Hasimoto et al., 1975). Label scored as mononucleotides = 5000.

(c) Figures in parentheses from Birnstiel et al (1968). Molecular weight from combination of gel electrophoresis (Loening, 1968) and electron microscopic measurements (Wellauer & Dawid, 1974) =  $1.65 \times 10^6$ ,  $\approx$  4670 nucleotides. Approximately 80 nucleotides are in resolved alkali-stable material.

(d) Figures in parentheses from Jacobson et al (1974). Data for RNA molecular weight (Loening, 1968) and methylation (Klootwijk & Planta, 1974) are for yeast. They are assumed to be similar for Dictyostelium. Of approximately 3800 nucleotides 50 run as resolved, alkali stable material.

TABLE 3.6.

Nucleotide Compositions of 5.8S rRNA

<u>Species</u>	% Cp      Ap      Gp      Up + Up				Total mono- nucleotides(a)	Molar amount Up + Up
	Cp	Ap	Gp	Up + Up		
HeLa	30.9	18.7	28.3	22.1	155	34.3
L-cell	28.6	20.7	29.2	21.5	155	33.3
Xenopus	26.9	22.0	28.5	22.6	155	35.0
Dictyostelium	18.0	29.7	23.6	28.7	156	44.5

HeLa 5.8S RNA base composition is from 5 determinations using T<sub>2</sub> ribonuclease. Base compositions of other 5.8S rRNA's are the means of 3 determinations using T<sub>2</sub> hydrolysis.

(a) As indicated in the legend to table 3.1, HeLa 5.8S releases about 155 mononucleotides. Data on other vertebrate 5.8S RNA's suggest that these sequences are essentially identical to those of HeLa (Khan & Maden, 1976; Nazar et al. 1975). The sequence of yeast 5.8S RNA is 158 nucleotides long (Rubin, 1973), and is unmethylated. T<sub>1</sub> + Pancreatic fingerprints of Dictyostelium 5.8S RNA suggest that it is also unmethylated (Fig. 3.1). It has been assumed that Dictyostelium and yeast 5.8S RNA are the same size and that the residues from either ends of the molecule are the only ones which do not run as mononucleotides on electrophoresis.

Fig. 3.1

T<sub>1</sub> plus pancreatic RNase fingerprint of Dictyostelium 5.8S RNA. Conditions of digestion and electrophoresis are described in section 2.4. The approximate positions expected for two methylated products (Um-Gp, Gm-Cp) found in vertebrate 5.8S RNA's (Khan & Maden, 1976) are marked. These spots are absent from this fingerprint.

$U_m G_p \rightarrow$

$G_m C_p \rightarrow$

$G_p$

$C_p$

$U_p$

are available on the rRNA of another fungus, yeast. The molecular weight and degree of methylation of Dictyostelium high molecular weight rRNA have been assumed to be similar to those for yeast. Dictyostelium 5.8S rRNA was fingerprinted after  $T_1$  plus pancreatic RNase digestion (Fig. 3.1). Two methylated sequences which are found in 5.8S rRNA of vertebrates (Khan & Maden, 1976) are absent. These methylations are also absent from yeast 5.8S rRNA. It seems likely that the rRNA's of yeast and dictyostelium will share other common features.

### 3.2.2 Pseudouridine Content

The percentage of  $\psi$ p in  $\psi$ p plus  $\psi$ p was quantitated for the ribosomal RNA's of L-cells, *Xenopus laevis* and Dictyostelium discoidium. The two-dimensional paper chromatographic method described for HeLa cell rRNA was used. Approximate numbers of  $\psi$ p residues present in each rRNA were calculated. Tables 3.7, 3.8 and 3.9 show the results of this investigation.

The standard deviation of the results for 5.8S RNA's is markedly higher than those for the high molecular weight rRNA's. As suggested for HeLa 5.8S RNA this may be due to errors in scintillation counting. In the case of dictyostelium 5.8S rRNA the amount of radioactivity present in  $\psi$ p was insufficient to cause more than faint blackening of X-ray film even after autoradiography for several days. Because of this, the position of  $\psi$ p on the

chromatogram was estimated from the position of the strong Up spot.

Estimated numbers of Up residues in rRNA's of L-cells and HeLa cells are very similar. This finding is not unexpected for the rRNA's of two closely related species. The figures obtained for *Xenopus laevis* are somewhat higher than for the other vertebrate species, particularly that for 18S rRNA. This finding was somewhat unexpected but the results of determinations from two independent rRNA preparations were in good agreement. The pseudouridine content of RNA's from *Dictyostelium* is markedly lower than that for vertebrate rRNA despite the fact that *Dictyostelium* rRNA has a high A + U content. It is obvious that the number of pseudouridines is not simply a constant fraction of the number of uridine residues present. The 18S, 28S and 5.8S RNA's with the most uridines (*Dictyostelium*) have the fewest pseudouridines.

Direct sequence data on the pseudouridine content of several 5.8S rRNA's are available (Nazar et al, 1975a; Khan & Maden, 1976; Rubin, 1973). Pseudouridine has been identified at two sites in all the vertebrate 5.8S rRNA's examined. Analysis of oligonucleotides from T<sub>1</sub> fingerprints of HeLa 5.8S has also shown the presence of a small amount of Up in the unresolved mixture of products 23 and 23a (Maden & Robertson, 1974; and B. E. H. Maden unpublished observation). The data therefore suggest that the 5.8S sequence contains pseudouridine at three sites. This is in agreement with the results shown in table 3.9. The sequence of yeast 5.8S RNA (Rubin, 1973) contains only one pseudouridine



TABLE 3.7

Pseudouridine Content of 18 S rRNA

<u>Species</u>	$\frac{\Psi_p}{U_p + \Psi_p} \%$	<u>Molar Amount</u>	Plus (b)	Total $\Psi$	No. of Determinations
HeLa	8.59 $\pm$ 0.14	36 $\pm$ 1	1	37 $\pm$ 1	5
L-cell	8.81 $\pm$ 0.22	34 $\pm$ 1	1 ?	35 $\pm$ 1	3
Xenopus	10.94 $\pm$ 0.56	50 $\pm$ 3	1	51 $\pm$ 3	5
Dictyostelium	1.76 $\pm$ 0.26	10 $\pm$ 1	1 ?	11 $\pm$ 1	3
Blank (a) 0.13					

$\Psi_p$  content was determined as described in the text.

(a) The blank value was obtained by rerunning previously purified  $U_p$ . This value has been subtracted from the other values.

(b) The additional  $\Psi_p$  is that found in a hypermodified nucleotide in 18 S rRNA of HeLa, Chick, Xenopus and yeast. This oligonucleotide is probably also present in L-cells and Dictyostelium.

TABLE 3. 8.

Pseudouridine Content of 28 S rRNA

<u>Species</u>	$\frac{\psi p}{Up + \psi p} \%$	Molar Amount	Plus (b)	Total $\psi p$	Number of Determinations
HeLa	7.29 $\pm$ 0.23	58 $\pm$ 2	2	60 $\pm$ 2	5
L-Cell	7.82 $\pm$ 0.17	57 $\pm$ 1	2	59 $\pm$ 1	3
Xenopus	8.03 $\pm$ 0.36	63 $\pm$ 3	2	65 $\pm$ 3	5
Dictyostelium	2.69 $\pm$ 0.19	29 $\pm$ 2	1 ?	30 $\pm$ 2	3
Blank (a) 0.13					

$\psi p$  content was determined as described in the text.

(a) The blank value was obtained by rerunning previously purified Up. This value has been subtracted from the other values.

(b) Two additional  $\psi p$  residues have been found in alkali-stable products in HeLa, Xenopus (Khan & Maden, 1976) and mouse (Hashimoto et al, 1975). Yeast 28S rRNA contains one  $\psi p$  residue in alkali stable linkage (Klootwijk & Planta, 1974). It is assumed that rRNA of Dictyostelium contains the same sequence.

TABLE 3.9

Pseudouridine Content of 5.8S rRNA

<u>Species</u>	$\frac{\Psi p}{Up + \Psi p} \%$	Molar Amount	(previous data) (b)	Number of Determinations
HeLa	$7.75 \pm 0.85$	$2.7 \pm 0.2$	( $\leq 2$ )	5
L-Cell	$8.92 \pm 0.54$	$3.0 \pm 0.2$	( $\leq 2$ )	3
Xenopus	$9.07 \pm 0.79$	$3.2 \pm 0.3$	( $\leq 2$ )	3
Dictyostelium	$3.46 \pm 1.02$	$1.5 \pm 0.4$	(1?)	3

Blank (a) 0.13

$\Psi p$  content was determined as described in the text

(a) The blank value was obtained by rerunning previously purified Up. This value has been subtracted from the other values.

(b) Two  $\Psi p$  residues have been identified in the 5.8 S sequences of several vertebrates (Nazar et al; Khan & Maden, 1976). A single  $\Psi p$  residue is present in the sequence of yeast 5.8S RNA (Rubin, 1973). It seems possible that Dictyostelium 5.8S RNA contains a similar sequence.

residue. Results shown here suggest that Dictyostelium 5.8S RNA probably also contains only one  $\Psi$ p.

### 3.2.3. Pseudouridine and 2'-O-Methylation: interspecies comparison.

Table 3.10 shows the number of 2'-O-methyl groups found in the rRNA of several organisms. The overall numbers of methylations in the vertebrate RNA's are very similar. As mentioned above, the number of methyl groups in the major rRNA's of Dictyostelium is unknown. 5.8S RNA of Dictyostelium, like that of yeast, is apparently unmethylated at two sites where methylation occurs in vertebrate 5.8S RNA's.

The numerical correlation between pseudouridine and 2'-O-methyl groups is extremely close for 18S RNA's of HeLa and L-cells. Rather more  $\Psi$ p residues were found in *Xenopus laevis* 18S RNA than in either of the mammalian species. In this case 2'-O-methylation and pseudouridylation do not appear to be closely numerically related. The number of 2'-O-methyl groups in Dictyostelium 18S rRNA is probably similar to that found in the related species, yeast (18 2'-O-methyl groups) (Klootwijk & Planta, 1974). If this is so there is, at least, an approximate numerical correlation between pseudouridine and 2'-O-methylation.

The correlation between 2'-O-methyl groups and  $\Psi$ p residues appears to be quite close for all the 28S RNA's studied. In the case of 5.8S RNA it appears that the two

TABLE 3.10

The number of 2'-O-Methyl Groups in rRNA

<u>Species</u>	<u>18 S</u>	<u>28 S</u>	<u>5.8 S</u>
HeLa (a) + (b)	38	62	1.2
L- cells (c)	35	65	$\geq 1$ ?
Xenopus (b)	33	66	1.4
Yeast (d)	18	37	0

(a) From Maden & Salim (1974)

(b) B. E. H. Maden, unpublished observations.

(c) From Hashimoto et al (1975). One methylated dinucleotide (GmC) was observed which was present when 28S was extracted cold but not when it was heated. This dinucleotide is probably from the 5.8 S sequence.

(d) From Klotwijk & Planta (1974) (28 S & 18S) and Rubin (1973) (5.8

secondary modifications occur in similar, although not equal numbers. For example, Dictyostelium 5.8S RNA is unmethylated but contains one or two pseudouridines. It may be that the two modifications are numerically related although not necessarily on a "one-to-one" basis.

### 3.3. Pseudouridine in HeLa Ribosomal Precursor rRNA

It has been conclusively shown that 32S rpre RNA contains the same number of methyl groups as 28S rRNA and that the number of methylations in 45S rpre RNA is equal to the sum of 2'-O-methyl groups in 18S and 28S RNA's. This indicates that the transcribed spacer regions of rpre RNA contain no methyl groups (Maden & Salim, 1974, discussed in Chapter 1). In view of the numerical correlation between 2'-O-methylation and pseudouridylic acid in HeLa rRNA it was thought to be of interest to investigate the level of pseudouridine in r pre RNA, with a view to determining whether it is also confined to the ribosomal sequences.

#### 3.3.1. Base Composition

Table 3.11 shows the base composition data for HeLa cell r pre RNA obtained from several  $T_2$  hydrolysates of both 45S and 32S RNA's. The results obtained for 45S RNA show a G + C content of about 65%. Two groups (Willemis et al., 1968; Jeanteur et al, 1968) purifying 45S r pre RNA by different methods obtained figures of around 70% G + C content. The discrepancy between the data of these groups and those described here is most

TABLE 3.11

Nucleotide Composition data on HeLa cell r pre RNA

	percentage				Total mono-nucleotides (a)	Molar amount Up + $\Psi$ p
	Cp	Ap	Gp	Up		
45 S	32.1	16.0	33.2	18.7	13,270	2480
32 S	32.6	14.3	37.6	15.5	6,500	1010

The results for both 45S and 32S RNA's are the means of 5 determinations using  $T_2$  ribonuclease.

(a) The total numbers of mononucleotides per RNA were estimated from published molecular weight values and sequence data as follows. 45 S RNA:- molecular weight from electron microscopic contour lengths (Wellauer & Dawid, 1973),  $4.7 \times 10^6$ ; from sedimentation analysis (Jeanteur, Amaldi & Attardi, 1968),  $4.6 \times 10^6$ ; from polyacrylamide gel electrophoresis (Weinberg & Penman, 1970),  $4.1 \times 10^6$ ; mean 45 S value  $4.47 \times 10^6 = 13410$  nucleotides. Of these approximately 140 would be released as products migrating differently from mononucleotides on electrophoresis (see legend to table 1). Therefore the label scored as mononucleotides in the hydrolysates = 13,270 nucleotides, of which Up +  $\Psi$ p = 18.7%, of 2480. 32 S RNA:- molecular weight from electron microscopy (Wellauer & Dawid, 1973),  $2.3 \times 10^6$ ; sedimentation analysis (Jeanteur et al, 1968),  $2.2 \times 10^6$ ; polyacrylamide gels (Weinberg & Penman, 1970),  $2.1 \times 10^6$ ; mean  $2.2 \times 10^6 \approx 6,600$  nucleotides. Subtract  $\sim 80$  for resolved material from 28S rRNA and 5.8 S rRNA (table 3.1 approximately 6520 migrating as mononucleotides.

likely to be due to contamination of 45S preparations with HnRNA since there may be some overlap between these two RNA types on sucrose gradients.

$^{32}\text{P}$  labelled Hn RNA was prepared as described by Fraser et al (1973). The base composition of the RNA was analysed after  $\text{T}_2$  hydrolysis. Mean values from three determinations gave a G + C content of 41-8%. The presence of Hn RNA in 45S RNA preparations would, therefore, be expected to give a lower G + C percentage than the figure for pure 45S r pre RNA. This is in accordance with the results in table 3.11.

The base composition data for 32S r pre RNA are in reasonable agreement with previously published findings (Jeanteur et al, 1968, Willems et al, 1968 ). This suggests that 32S preparations are substantially pure.

The approximate molar amounts of Up +  $\Psi$ p in 32S and 45S rpre RNA were calculated using base composition data from table 3.11 and published molecular weights.

### 3.3.2. Pseudouridine Content

Quantitation of  $\Psi$ p in HeLa r pre RNA was carried out using the method described for HeLa rRNA. In this case only  $\text{T}_2$  hydrolysis was used. Table 3.12 shows the results of this analysis.

The estimate of total numbers of  $\Psi$ p in 32S RNA is somewhat higher than that for 28S RNA. It is also rather higher than that calculated from data of Jeanteur et al (1968). These results suggest that the transcribed spacer material, which is lost



TABLE 3.12

Pseudouridylate content of HeLa cells rRNA.

	$\frac{\psi_p}{\text{Up} + \psi_p} \%$	Molar Amount	(previous data) (b)	Plus (c)	Total	Number of Determination
45 S	$2.85 \pm 0.17$	$71 \pm 4$	( $\simeq 117$ )	3	$74 \pm 4$	5
32 S	$6.87 \pm 0.12$	$69 \pm 1$	( $\simeq 59$ )	2	$71 \pm 1$	5
Blank (a)	0.13	-	-	-	-	3

(a) The blank value was obtained by rerunning previously purified Up. This value has been subtracted from other values.

(b) These values were calculated from data of Jeanteur et al (1968), who recorded a base composition of 0.85%  $\psi_p$  for 45 S RNA and 0.90%  $\psi_p$  for 32 S RNA.

(c) Two pseudouridines from 28S RNA and one from 18 S RNA are not liberated as free  $\psi_p$  as indicated in the text.

when 32S r pre RNA is converted to 28S rRNA, contains a few  $\psi$ p residues. This finding does not favour an exact topographical correlation between  $\psi$ p and 2'-O-methylation in 32S RNA since, as already discussed, the transcribed spacer contains no methyl groups.

Robertson & Maden (1973) studied the long unique oligonucleotides found in  $T_1$  fingerprints of 32S RNA but absent from fingerprints of 28S RNA. No  $\psi$ p residues were detected in these characteristic products. However, only some 20% of uridine residues in the 32S transcribed spacer could be accounted for by these oligonucleotides. The remainder occur in short, multiply occurring sequences or in long sequences which were incompletely resolved. Nothing is known about the pseudouridine content of these products. Results described in this thesis may not therefore conflict with the findings of Robertson & Maden.

The estimated number of pseudouridine residues in 45S r pre RNA is considerably less than the sum of the numbers determined for 28S and 18S rRNA's. However, calculations from the data of Jeanteur et al suggest that approximately 117  $\psi$ p residues are present in 45S RNA, a figure rather greater than the sum of  $\psi$ p residues in 28S and 18S RNA's.

One possible reason for a relatively low  $\psi$ p content for 45S RNA might be contamination with Hn RNA. Hn RNA was analysed for pseudouridine content using the two dimensional

chromatographic procedure already described. After subtraction of the blank value, less than 0.1% of Up plus  $\psi$ p was found in the  $\psi$ p area of the chromatogram. Thus the percentage of  $\psi$ p in Up plus  $\psi$ p in "45 S RNA" is probably an underestimate because of the contribution of contaminating Hn RNA to the total radioactivity.

Another possible reason why estimates of  $\psi$ p in 45S RNA may be too low is that modification of Up to  $\psi$ p may occur sometime after synthesis of the RNA. For instance, if the lifetime of 45S RNA is 20 minutes and modification to  $\psi$ p occurs, on average, 5 minutes after synthesis, then observed  $\psi$ p content would only be 75% of the final amount in rRNA.

Further analysis of individual sequences in 32S and 45S RNA, as described for rRNA (next chapter), would be desirable. However, the extremely large size of the r pre RNA molecules makes this technically very difficult.

## Chapter 4

### Fingerprinting Studies on Pseudouridine in HeLa rRNA.

Having determined the pseudouridine content of whole ribosomal RNA (Chapter 3), it seemed to be of interest to investigate pseudouridine in individual oligonucleotides from rRNA. This study was undertaken with two related questions in mind. These questions were:- (1) Does the conversion of Up to  $\psi$ p occur at specific sites in rRNA, or is pseudouridine spread randomly through the uridine containing sequences?

(2) Do all the molecules of a particular rRNA contain pseudouridine in a given sequence, i. e. is conversion of Up to  $\psi$ p an "all-or none" event?

In addition experiments were carried out to investigate the sequence of oligonucleotides containing  $\psi$ p. This study was, however, secondary to the two main questions and only partial sequence analysis was attempted.

Separation of oligonucleotides from rRNA was carried out by use of the RNA fingerprinting techniques of Sanger et al (1965) (Described in Methods). Two fingerprinting systems were used in this study. Optimum separation of oligonucleotides containing one uridine was achieved by digestion of rRNA with  $T_1$  RNase and separation for 40 hours in the second dimension of electrophoresis ("long  $T_1$ " system).

Good resolution of oligonucleotides containing two or three uridines was obtained by separation of the products of  $T_1$  RNase plus alkaline phosphate digestion for 40 hours in the second dimension ( $T_1$  plus phosphatase system).

#### 4.1 Long $T_1$ Fingerprints of HeLa 18S rRNA

Fig 4.1 (a) shows an autoradiograph of a typical long  $T_1$  separation of oligonucleotides from  $^{32}\text{P}$  labelled HeLa 18S rRNA. Many, but not all, of the spots in the "one U" graticule are completely resolved from their neighbours. In cases where two spots were not fully resolved it was sometimes necessary to estimate a dividing line between them when excising spots for further investigation. Uncertainty in dividing spots may give rise to contamination of one oligonucleotide by another. An example of this is spot 155. This spot forms the fastest running part of a poorly resolved, elongated spot which contains three different nucleotide sequences. No visual separation of the spot is possible and a somewhat arbitrary division had to be made.

#### Analysis of Products

Spots from the one U graticule of long  $T_1$  fingerprints were excised and assayed for radioactivity. Oligonucleotides were eluted from DEAE paper and analysed for base composition and pseudouridine content.

Complete hydrolysis of oligonucleotides for base composition

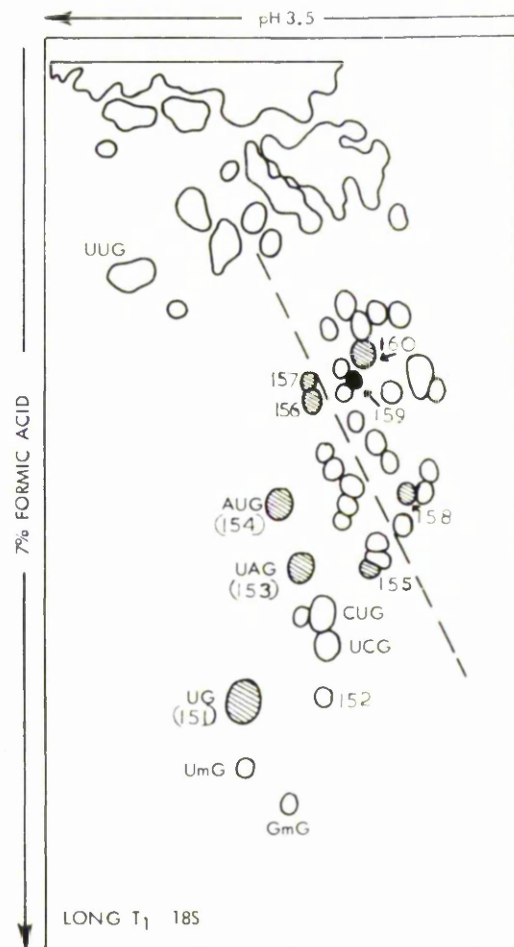
Fig. 4.1

(a) Long  $T_1$  RNase fingerprint of  $^{32}\text{P}$ -labelled HeLa 18S RNA. Separation conditions are described in Section 2.4.

(b) Key. Products above and to the right of the interrupted line are penta- and larger oligonucleotides, most of which occur once per molecule. However, not all the unique products are resolved from each other. Open circles, no pseudouridylate, closed circles, pseudouridylate content suggests one pseudouridylate in a unique sequence. Cross-hatched circles, multiple yield products or incompletely resolved mixtures with some pseudouridylate present.



a.



b.

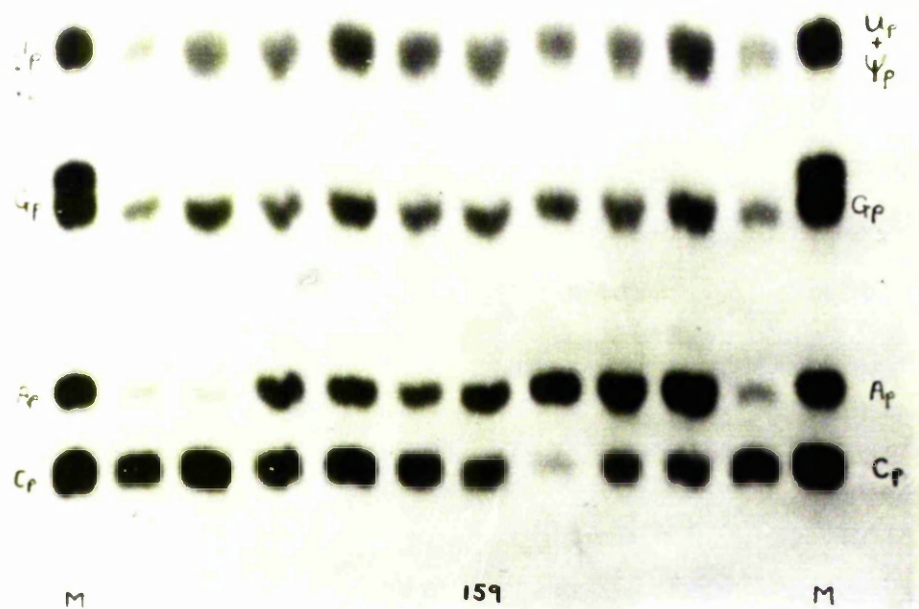
Fig. 4.2.

(a) pH 3.5 electrophoretic separation of alkali digests of spots from long  $T_1$  RNase fingerprints of 18S RNA. M is a mixture of mononucleotides made by complete digestion of  $^{32}\text{P}$ -labelled cytoplasmic RNA.

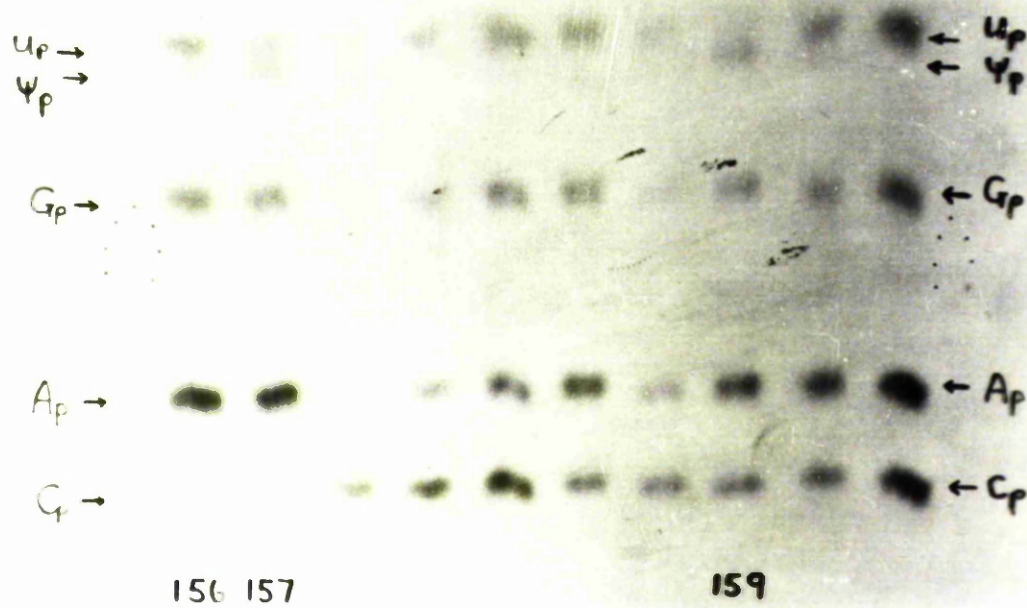
(b) pH 3.5 electrophoretic separation of  $T_2$  digests of spots from long  $T_1$  RNase fingerprints of 18S RNA.  $\psi$ p is slightly separated from Up in this system, as shown by spots 157 and 159. Spot 157 contains approximately equal amounts of Up and  $\psi$ p whilst spot 159 contains very little Up.



a.



b.



or pseudouridine analysis was carried out with alkali or  $T_2$  ribonuclease. Resolution of mononucleotides after pH 3.5 electrophoresis was sharper for  $T_2$  digests than for alkali digests. In the case of spots containing pseudouridine it is sometimes possible to detect  $\psi p$  as a band of material running slightly behind Up in separations of  $T_2$  digests (Fig 4.2 - note spots 157 and 159). Resolution of mononucleotides from alkali digests is insufficient for this type of preliminary detection of  $\psi p$ . Maden and Salim (1974) assayed radioactivity in methylated products in several 18S fingerprints. A number of spots were found to be approximately equally labelled and a mean value of "counts per minute per methyl group" was determined. Individual molar yields were expressed relative to these means. On this basis the molar yields of the dinucleotides GmG and UmG were very close to unity.

The chain length of oligonucleotides was calculated from base composition data. Knowing the number of nucleotides present and the total radioactivity in each spot, it was possible to determine the counts per minute (c.p.m.) per nucleotide in each case. Radioactivity in the methylated dinucleotides UmG and GmG was also assayed. As both these products are known to be unimolar it was possible to determine a mean c.p.m. per mole of nucleotide. The molar frequency of an oligonucleotide was then determined by dividing c.p.m.

Fig. 4.3

Chromatographic separation of Up and  $\psi$ p.

Oligonucleotides from 18S long  $T_1$  RNase fingerprints

were completely hydrolysed with  $T_2$  RNase and

mononucleotides separated by electrophoresis at pH 3.5.

The Up plus  $\psi$ p graticule was subjected to chromatography

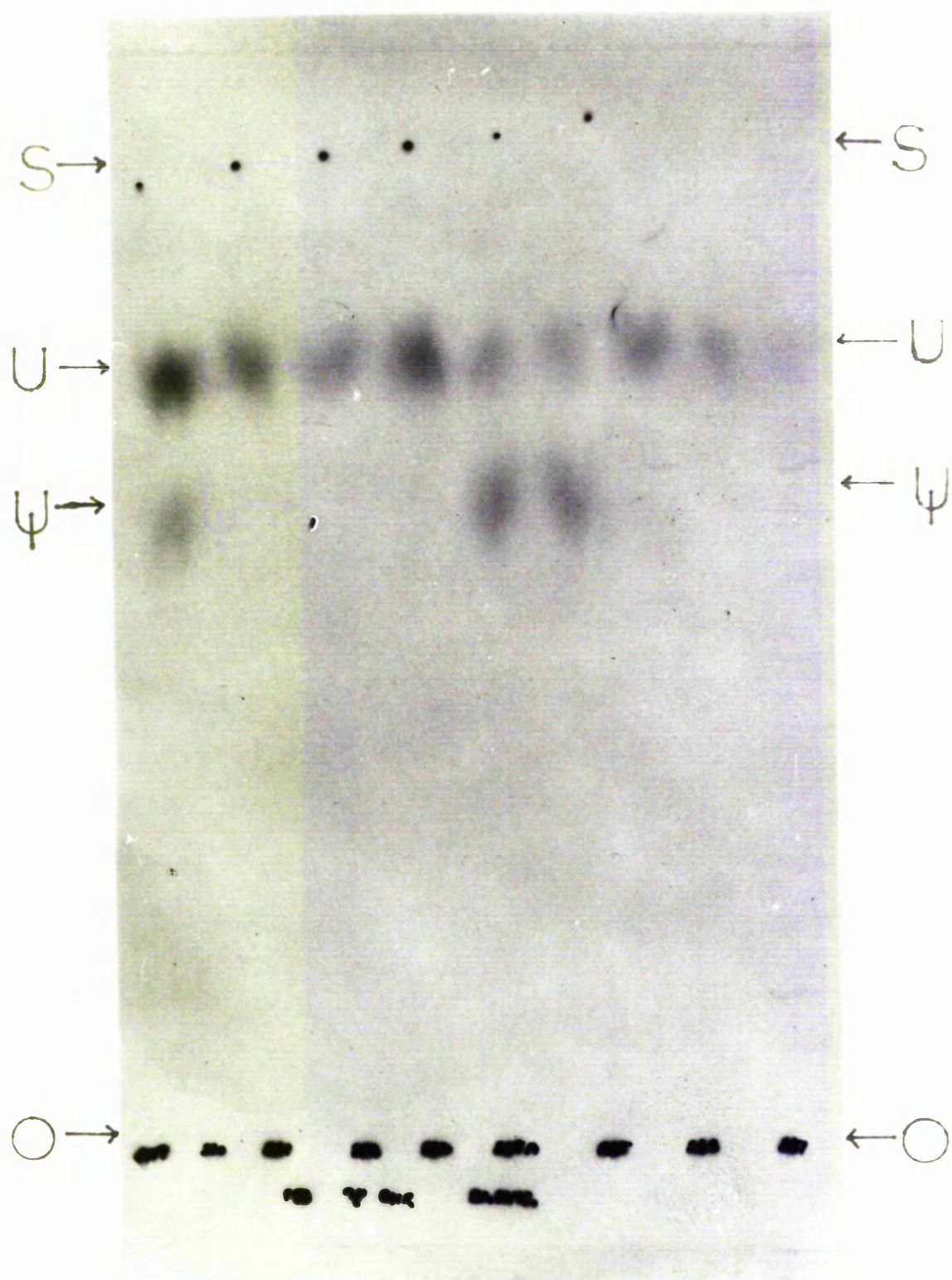
with propan-2-ol/HCl/H<sub>2</sub>O solvent.

O-marks the position of material at the start of

chromatography.

S - marks the position of the solvent front at the

termination of chromatography.



per nucleotide by c. p. m. per mole of nucleotide.

Thus:- GmG - x c. p. m.

UmG - y c. p. m.

Therefore mean c. p. m. per mole of mononucleotide =  $\frac{x + y}{4}$

Spot A (n nucleotides) - a c. p. m.

c. p. m. per nucleotide =  $\frac{a}{n}$

Molarity of Spot A =  $\frac{\frac{a}{n}}{\frac{x + y}{4}}$

All oligonucleotides in the one U graticule were analysed for the presence of  $\psi$ p. This was achieved by complete hydrolysis, pH 3.5 electrophoresis, and one-dimensional chromatography of the "Up +  $\psi$ p" electrophoretic band (Fig 4.3). Some trailing of Up into the  $\psi$ p region of the chromatogram was detected. The amount of trailing was variable but was sometimes as much as 5% of counts in previously purified Up. In certain high frequency spots such as UGp (Spot 151), a visible  $\psi$ p spot was observed although only a low percentage of the total radioactivity was found in this area of the chromatogram. It was considered that pseudouridine was present in a spot if either greater than 5% of Up +  $\psi$ p counts were detected in the  $\psi$ p area or if a visible  $\psi$ p spot was seen on chromatograms.

Fig. 4.1b shows a diagram of a long  $T_1$  separation. Spots which contain reproducibly detectable amounts of  $\psi$ p are shaded. Data

on the molar yield, base composition and pseudouridine content of these spots are shown in Table 4.1. Most of the  $\Psi$ p containing spots occur more than once per 18S molecule. In no case is Up completely absent. Spot 159 is a unimolar spot and contains a high  $\Psi$ p percentage. The Up detected may be due to incomplete resolution of this spot from two adjacent spots which do not contain  $\Psi$ p.

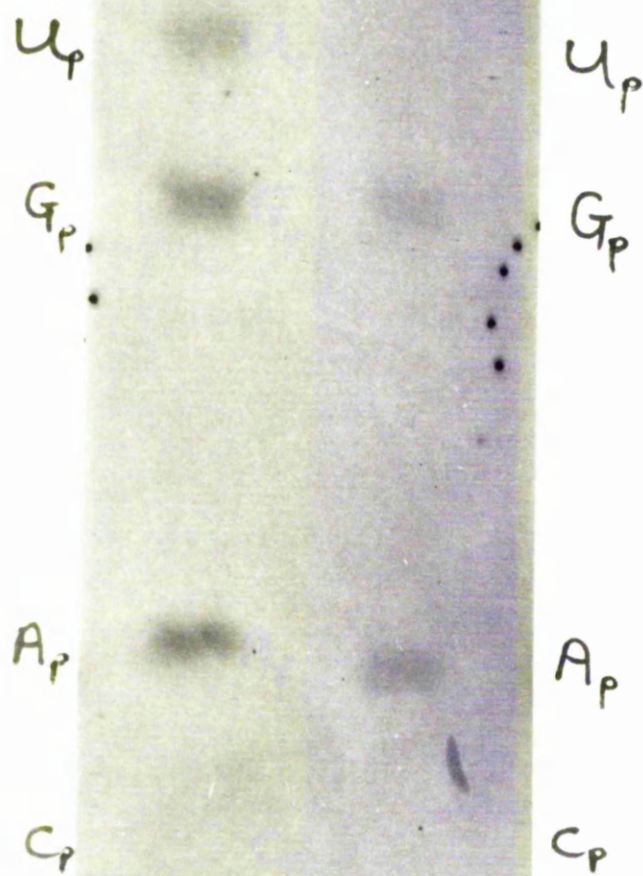
$T_2$  digestion of spot 152 suggests that it has a base composition of UAG. However, when this spot is digested with alkali no Up is detected (Fig. 4.4). This suggests that spot 152 may contain dihydrouridine which is unstable in alkali but runs as uridine in enzymic digests (Brownlee, 1972). The molar yield of this spot is consistently less than one.

Electrophoresis of mononucleotides on Whatman 52 paper shows that Up has a slightly greater mobility than  $\Psi$ p at pH 3.5. Similarly,  $T_1$  plus pancreatic ribonuclease fingerprints of HeLa 18S rRNA (Fig. 4.5) show that short oligonucleotides terminated by  $\Psi$ p are separated from homologous sequences terminated by Up. Differences in mobility between longer homologous sequences, not terminated by Up or  $\Psi$ p, seem not to be as great as this. Homologous sequences differing only in the presence or absence of a single pseudouridine are not completely resolved in fingerprints. A pseudouridine-containing sequence may, however, have sufficient difference in mobility from its uridine containing homologue so that it overlaps with an adjacent spot. An example of this is provided

Fig. 4.4

pH 3.5 electrophoretic separation of the products of hydrolysis of spot 152 from long  $T_1$  RNase fingerprints of HeLa 18S RNA. Material in  $T_2$  hydrolysates which has the same electrophoretic mobility as marker Up is absent from alkali hydrolysates. This suggests that spot 152 contains dihydrouridine.





$T_2$  Alkali  
Spot 152



Fig. 4.5

$T_1$  plus pancreatic RNase fingerprint of  $^{32}\text{P}$ -labelled HeLa 18 S RNA. Conditions of digestion and separation are described in section 2.4. Short sequences containing pseudouridine have slightly different mobility from homologous sequences containing uridine.

T. + Marc 30.1.75

TABLE 4.1

Pseudouridine Containing Oligonucleotides from 18S Long T<sub>1</sub> Fingerprints.

<u>Spot 151</u>	Base Composition	-	U <sub>1</sub>	G <sub>1</sub>
	Mean Molar Yield	=	34.1 <sup>+</sup> <sub>-</sub>	2.9 (5 determinations)
	$\frac{\psi p}{U_p + \psi p} \%$	=	3.5 <sup>+</sup> <sub>-</sub>	0.5 (4 determinations)

Comment    Approximately one  $\psi p$  per molecule of 18S RNA  
within the sequence, G -  $\psi$  - G

<u>Spot 152</u>	Base Composition	-	A <sub>1</sub>	U <sub>1</sub> *	G <sub>1</sub>
	Mean Molar Yield	=	0.5 <sup>+</sup> <sub>-</sub>	0.1	(4)

\* T<sub>2</sub> digest shows Up spot after electrophoresis  
Alkali digest does not (see fig. 4.4)

Comment    This oligonucleotide occurs in submolar yield. It  
probably contains dihydrouridine. This is unstable in  
alkali and breaks down to yield  $\beta$ -ureidopropionic acid  
N - ribotide (Brownlee, 1972). Dihyrouridine is  
stable towards ribonuclease T<sub>2</sub>.

<u>Spot 153</u>	Base Composition	-	A <sub>1</sub>	U <sub>1</sub>	G <sub>1</sub>
	Pancreatic digestion Products		U,	AG	
	Sequence		UAG <sub>p</sub>		
	Mean Molar Yield	=	5.4 <sup>+</sup> <sub>-</sub>	1.4	(5)
	$\frac{\psi p}{U_p + \psi p} \%$	=	12.8 <sup>+</sup> <sub>-</sub>	1.1	(3)

Comment    Multiple yield spot.  $\psi p \leq 1$  per 18S molecule in  
this sequence.

---

Spot 154Base Composition - A<sub>1</sub> U<sub>1</sub> G<sub>1</sub>

Pancreatic Digestion Products - AU, G

Sequence - AUGp

Mean Molar Yield = 12.4  $\pm$  2.5 (5)
$$\frac{\psi p}{U_p + \psi p} \% = 8.1 \pm 0.9 (4)$$

Comment These values imply that G-A-U-G occurs eleven times and G - A -  $\psi$  - G once per molecule.

---

Spot 155Base Composition - U<sub>1</sub> C<sub>2</sub> G<sub>1</sub>Mean Molar Yield = 2.1  $\pm$  0.6 (5)
$$\frac{\psi p}{U_p + \psi p} \% = 28.5 \pm 1.5 (3)$$

Comment Incompletely resolved from neighbouring spot. Molar yield and  $\psi p$  content rather variable because of this.  
 $\psi p \leq 1$  per molecule in this sequence.

---

Spot 156Base Composition - U<sub>1</sub> A<sub>2</sub> G<sub>1</sub>

Pancreatic Digestion Products - AAU, G

Sequence - AAUGp

Mean Molar Yield = 3.4  $\pm$  0.6 (5)
$$\frac{\psi p}{U_p + \psi p} \% = 50.2 \pm 1.3 (3)$$

Comment Sequence G-A-A- $\psi$  - G probably occurs twice per 18S molecule. Overlap with spot 157 makes quantitation difficult.

---

---

Spot 157Base Composition - U<sub>1</sub> A<sub>2</sub> G<sub>1</sub>

Pancreatic Digestion Products - AU, AG

Sequence - AUAGp

Mean Molar Yield = 1.3  $\pm$  0.2 (5)

$$\frac{\psi_p}{U_p + \psi_p} \% = 30.6 \pm 2.4 (3)$$

Comment Overlap with spot 156.  $\psi_p$  apparently in submolar yield.  
This may be due to spillover of  $\psi_p$  from 156.

---

Spot 158Base Composition - U<sub>1</sub> C<sub>3</sub> G<sub>1</sub>Mean Molar Yield = 1.7  $\pm$  0.2 (6)

$$\frac{\psi_p}{U_p + \psi_p} \% = 25.5 \pm 2.3 (3)$$

Comment  $\psi_p$  certainly present but appears to be in submolar yield.  
Some overlap with adjacent spot makes quantitation uncertain.

---

Spot 159Base Composition - U<sub>1</sub> A<sub>2</sub> C<sub>1-2</sub> G<sub>1</sub>

Pancreatic Digestion Products - C, AU, AG

Mean Molar Yield = 1.1  $\pm$  0.3 (5)

$$\frac{\psi_p}{U_p + \psi_p} \% = 75.5 \pm 3.1 (3)$$

Comment Unique spot. High  $\psi_p$  percentage but some U<sub>p</sub> present.  
May be contamination from several adjacent spots.  
Seems probable that  $\psi_p$  occurs once per 18S molecule  
within this oligonucleotide.

---

---

Spot 160

Base Composition      -    $U_1$     $A_1$     $C_2$     $G_1$   
Mean Molar Yield      =    $2.0 \pm$     $0.6$    (5)

$$\frac{\psi_p}{U_p + \psi_p} \% = 46.7 \pm 1.5 (3)$$

Comment    $U_p$  containing and  $\psi_p$  containing sequences each  
                 occur once per 18S molecule.

---

by spots 156 and 157 which are incompletely separated in fingerprints. Pseudouridine is detected in both spots. The amount found in spot 157 at first suggests a submolar conversion from Up. However, this may be because Up-containing sequences homologous to sequences in spot 156 have electrophoretic mobility intermediate between spots 156 and 157. When these spots are cut out part of the Up containing material would be excised with each of the Up containing sequences.

#### 4.2 T<sub>1</sub> plus Phosphatase Fingerprints of HeLa 18S rRNA.

Fig. 4.6 (a) shows a T<sub>1</sub> plus phosphatase fingerprint of <sup>32</sup>P HeLa 18S rRNA. In this type of separation oligonucleotides containing only one Up have been run off the fingerprint in order to give maximum separation in the 'two U' and 'three U' graticules. Molar yields of oligonucleotides in these graticules were quantitated relative to two unimolar methylated sequences (M 41 and M47 of Maden & Salim, 1974). Because of the removal of the terminal phosphate group by alkaline phosphatase, only (n - 1) phosphate groups will contribute to radioactivity estimation of an oligonucleotide containing n nucleotides.

Thus:- Spot B (n nucleotides) - b. c. p. m.

$$\text{c. p. m. per nucleotide} = \frac{b}{n-1}$$

Base compositions of the spots were determined and their pseudouridine contents analysed. Limited sequence analysis was carried out by complete digestion of spots with

Fig. 4.6

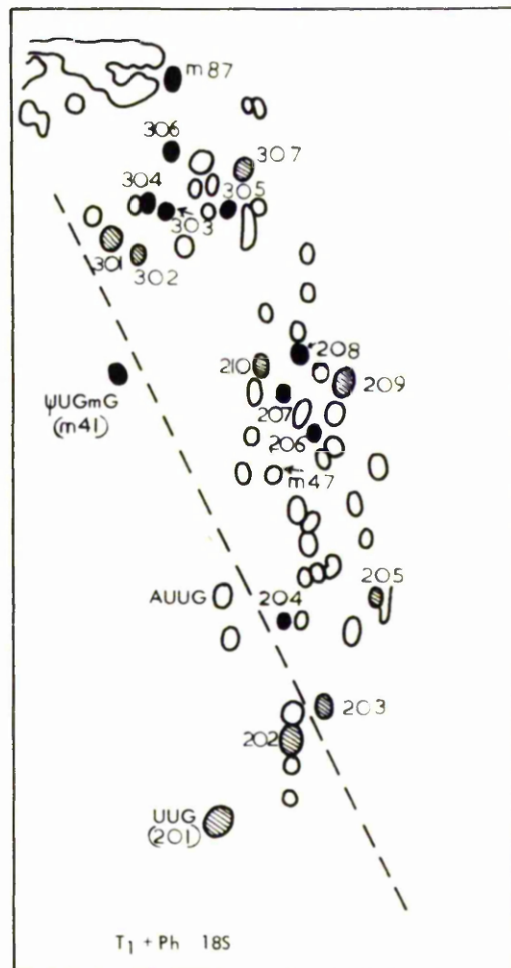
(a)  $T_1$  RNase plus alkaline phosphatase fingerprint of  $^{32}\text{P}$ -labelled HeLa 18 S RNA. Conditions of digestion and separation are described in section 2.4

(b) Key. Products above and to the right of the interrupted line are penta- and longer oligonucleotides, most of which occur once per molecule. However, not all the unique products are resolved from each other. Open circles, no pseudouridylate, closed circles, pseudouridylate content suggests at least one pseudouridylate in a unique sequence. Cross-hatched circles, multiple yield products or incompletely resolved mixtures with pseudouridylate present.





a.



b.

pancreatic ribonuclease or snake venom phosphodiesterase.

Table 4.2 shows the results of these investigations.

As was found in the one U graticule, there are a number of high yield spots which contain a low level of pseudouridine. However, many of the oligonucleotides occur at a molar frequency of only one or two. There are no spots in which the Up +  $\Psi$ p electrophoretic band contains only  $\Psi$ p. In a number of cases pseudouridine comprises 30-60% of Up +  $\Psi$ p. When such a sequence occurs more than once per molecule, it is not possible to determine unequivocally whether each oligonucleotide contains some  $\Psi$ p or whether all the  $\Psi$ p residues are in a single oligonucleotide. The mobility of a sequence containing a single  $\Psi$ p residue is not sufficiently different from that of a homologous sequence not containing  $\Psi$ p for resolution in fingerprints. However, a rather larger difference in mobility would be expected for a sequence containing two or three pseudouridines. If this is the case, at least partial resolution of a spot containing several pseudouridines from a homologous non- $\Psi$ p containing sequence would be expected. This appears not to be the case for many of the spots containing 30 - 60%  $\Psi$ p. For example, spot 210 is small and well separated from other spots. This nucleotide occurs twice per molecule and about half the "Up +  $\Psi$ p" band is  $\Psi$ p. It seems more likely that each oligonucleotide contains one  $\Psi$ p residue, than that one sequence contains two pseudouridines and the other one none,.

Pseudouridine -containing sequences in the 3U graticule may contain between one and three  $\psi$ p residues. Sequences containing two or three pseudouridines might be expected to be at least partly resolved from homologous "uridine only" sequences. Three of the four spots in this graticule which contain a high  $\psi$ p percentage are small and well resolved. This suggests that they probably contain only one  $\psi$ p in each copy of the oligonucleotide. The fourth spot (spot 307) containing a high %  $\psi$ p is rather broad. This is apparently due to the presence of a methylated sequence which may sometime be resolved from the other component of the spot. Results suggest that only the unmethylated sequence contains  $\psi$ p.

TABLE 4.2

Pseudouridine Containing Oligonucleotides from 18S T<sub>1</sub> plusPhosphatase Fingerprints2U Graticule

<u>Spot 201</u>	Base Composition	-	U <sub>2</sub> G <sub>1</sub>
	Mean Molar Yield	=	10.0 $\pm$ 2.2 (5 determinations)
	$\frac{\Psi p}{U p + \Psi p} \%$	=	13.5 $\pm$ 1.5 (3 determinations)
<u>Comment</u>	High yield spot. 2 - 4 $\Psi p$ residues per 18S molecule in either G-U- $\Psi$ -G or G- $\Psi$ -U-G		

<u>Spot 202</u>	Base Composition	-	U <sub>2</sub> C <sub>1</sub> G <sub>1</sub>
	Total Venom Digestion Products	-	U <sub>1</sub> C <sub>1</sub> G <sub>1</sub>
	Mean Molar Yield	=	5.3 $\pm$ 1.3 (5)
	$\frac{\Psi p}{U p + \Psi p} \%$	=	29.6 $\pm$ 1.8 (3)
<u>Comment</u>	Up at 5' end. 2-4 $\Psi p$ residues per 18S molecule in this oligonucleotide.		

<u>Spot 203</u>	Base Composition	-	U <sub>2</sub> C <sub>2</sub> G <sub>1</sub>
	Total Venom Products	-	U <sub>1</sub> C <sub>2</sub> G <sub>1</sub>
	Mean Molar Yield	=	4.0 $\pm$ 0.5 (4)
	$\frac{\Psi p}{U p + \Psi p} \%$	=	35.2 $\pm$ 2.1 (3)
<u>Comment</u>	Up or $\Psi p$ at 5' end. At least 3 $\Psi p$ residues per 18S molecule in this spot. Possibly one $\Psi p$ residue in each oligonucleotide.		

---

Spot 204

Base Composition -  $U_2 A_1 C_1 G_1$

Total Venom Products -  $U_1 C_1 G_1$

Pancreatic Digestion Products - U, AC

Mean Molar Yield =  $0.9 \pm 0.1$  (6)

$$\frac{\psi p}{U_p + \psi p} \% = 40.5 \pm 5.0 \text{ (2)}$$

Comment Sequence is ACUUG. Unique sequence probably containing one  $\psi p$  residue.

---

Spot 205

Base Composition -  $U_2 C_2 G_1$

Mean Molar Yield -  $1.3 \pm 0.2$  (4)

$$\frac{\psi p}{U_p + \psi p} \% = 33.0 \pm 3.3 \text{ (3)}$$

Comment Probably unique spot containing one  $\psi p$ . Quantitation effected by adjacent spot which consistently streaks in second dimension.

---

Spot 206

Base Composition -  $U_2 A_2 C_2 G_1$

Pancreatic Digestion products - AU, U, C.

Mean Molar Yield =  $0.8 \pm 0.1$  (6)

$$\frac{\psi p}{U_p + \psi p} \% = 34.2 \pm 1.9 \text{ (3)}$$

Comment Unique Spot. Probably contains one  $\psi p$ .

---

---

Spot 207

Base Composition - U<sub>2</sub> A<sub>3</sub> C<sub>1</sub> G<sub>1</sub>

Pancreatic Digestion Products - U, C, (A) AAG

Total Venom Products - A<sub>3</sub> C<sub>1</sub> U<sub>1</sub> G<sub>1</sub>

Mean Molar Yield = 1.2 <sup>±</sup> 0.3 (6)

$$\frac{\psi_p}{U_p + \psi_p} \% = 45.7 \pm 1.0 \quad (3)$$

Comment Unique spot containing one  $\psi_p$ . Venom digest suggests U at 5' end. If pancreatic product which runs slightly slower than marker Ap is ApApApG then oligonucleotide would be U(UC) AAAG. This is likely since pancreatic digest produces no AU or AC.

---

Spot 208

Base Composition U<sub>2</sub> C<sub>2</sub> A<sub>3</sub> G<sub>1</sub>

Pancreatic Digestion Products - C, U, AU, ApG, AC

Total Venom Products - C<sub>2</sub> A<sub>2</sub> U<sub>2</sub> G

Mean Molar Yield = 2.1 <sup>±</sup> 0.5 (7)

$$\frac{\psi_p}{U_p + \psi_p} \% = 53.4 \pm 0.6 \quad (3)$$

Comment 5-'Ap, 3' - AGp. 2  $\psi_p$  residues per molecule in this spot. Could be two oligonucleotides each containing one Up or one sequence containing two  $\psi_p$  residues. The former possibility seems more likely (discussed in text).

---

Spot 209

Base Composition A, U<sub>2</sub> C<sub>2-3</sub> G<sub>1</sub> (+ other \*)

Pancreatic Digestion Products - C, AC, U. (+ other \*)

Mean Molar Yield = 6.7 <sup>±</sup> 1.5 (4)

$$\frac{\psi_p}{U_p + \psi_p} = 15.7 \pm 4.3 \quad (3)$$

\* weakly labelled component migrating near marker Gp

Comment This spot probably includes one oligonucleotide which contains a single  $\psi_p$  residue. Methylated spot 49 of Maden & Salim (1974) which contains m<sup>7</sup>G may also be included.

---

---

Spot 210

Base Composition - U<sub>2</sub> A<sub>3</sub> C<sub>1</sub> G<sub>1</sub>

Pancreatic Digestion Products - C, ApG, U, AAU

Total venom products - U<sub>2</sub> A<sub>3</sub> G<sub>1</sub>

Mean Molar Yield = 1.8  $\pm$  2.4 (3)

$$\frac{\psi p}{U_p + \psi p} \% = 14.9 \pm 2.4 (3)$$

$\psi p$  in U spot, not in AAU

Comment Sequence is CUAAUAG or CAAUUAG and occurs twice per 18S molecule. A single  $\psi p$  residence probably occurs in one copy of this oligonucleotide.

---

3U Graticule

Spot 301

Base Composition - U<sub>3</sub> A<sub>1</sub> G<sub>1</sub>

Pancreatic Digestion Products - AU, U

Total Venom Products - A<sub>1</sub> U<sub>2</sub> G<sub>1</sub>

Mean Molar Yield = 3.6  $\pm$  0.7 (6)

$$\frac{\psi p}{U_p + \psi p} \% = 12.5 \pm 0.5 (3)$$

-  $\psi p$  in panc U product, not in AU.

Comment Sequence either UAUUG or UUAUG, 1 or 2  $\psi p$  residues per 18S molecule within this spot.

---

---

Spot 302

Base Composition -  $U_3 A_1 G_1$

Pancreatic Digestion Products - AU, U

Total Venom Products -  $U_3 G_1$

Mean Molar Yield =  $1.7 \pm 0.3$  (5)

$$\frac{\psi_p}{U_p + \psi_p} \% = 14.2 \pm 1.3 \text{ (3)}$$

-  $\psi_p$  in panc Up product

Comment

Sequence is AUUUG and probably occurs twice per 18S molecule. One of these oligonucleotides contains a single  $\psi_p$  residue.

---

Spot 303

Base Composition -  $U_3 A_1 C_1 G_1$

Pancreatic Digestion Products - C, AU, U

Total Venom Products -  $A_1 G_1 U_3$

Mean Molar Yield =  $1.5 \pm 0.2$  (6)

$$\frac{\psi_p}{U_p + \psi_p} \% = 28.2 \pm 2.9$$

-  $\psi_p$  in panc Up product

Comment

Sequence probably occurs twice per molecule (molar yield is always greater than 1) with one  $\psi_p$  in each copy.

---

Spot 304

Base Composition -  $U_3 A_1 C_1 G_1$

Pancreatic Digestion Products - C, AU, U

Total Venom Products -  $C_1 A_1 U_2 G_1$

Mean Molar Yield =  $1.97 \pm 0.5$  (6)

$$\frac{\psi_p}{U_p + \psi_p} \% = 24.9 \pm 1.9 \text{ (3)}$$

-  $\psi_p$  in panc AU product

Comment

Sequence which occurs twice per 18S molecule. Probably one Up residue in each copy.

---



---

Spot 305Base Composition -  $U_3 C_{2-3} G_1$ Mean Molar Yield =  $1.2 \pm 0.3$  (5)

$$\frac{\psi_p}{U_p + \psi_p} \% = 28.5 \pm 2.4 \quad (3)$$

Comment Unique sequence containing one  $\psi_p$  residue.

---

Spot 306Base Composition -  $U_3 A_3 G_1$ 

Pancreatic Digestion Products - AU (strong)

Mean Molar Yield =  $1.8 \pm 0.3$  (4)

$$\frac{\psi_p}{U_p + \psi_p} \% = 57.0 \pm 1.3 \quad (3)$$

Conclusion AUAUAUG occurs twice per 18S molecule. Total of 3-4  $\psi_p$  residues. It is not possible to assess the distribution of these in the oligonucleotides.

---

Spot 307

On some fingerprints this spot appears as a doublet. The two parts of this were designated 307A (faster running in 2nd D) and 307B.

Base Composition 307A -  $C_{1-2} A_{2-3} U_3 G_1$ 307B -  $C_{1-2} A_{2-3} U_3 G_1 + GmC$ Molar Yield - 2-3 (depends on exact base composition)  
(whole spot)

$$\frac{\psi_p}{U_p + \psi_p} \% \quad 307A = 52.3 \text{ (1 only)}$$

$$307 = 39.7 \pm 1.7 \quad (3)$$

Comment Part of this spot (307B) is M70 of Maden & Salim (1974). M. 70 probably does not contain  $\psi_p$ . It is likely that at least one oligonucleotide contained within 307A contains more than one  $\psi_p$  residue in its sequence.

---

Spot M87 (Maden & Salim, 1974)

Base Composition - contains A, U, G, AmU.

$$\frac{\Psi p}{U p + \Psi p} \% = 34.0 \pm 1.5 (2)$$

Comment      Long unique sequence which moves very little in  
second electrophoresis dimension. Contains at  
least one  $\Psi p$  residue.

---

#### 4.3 Long T<sub>1</sub> Fingerprints of HeLa 28S rRNA

Long T<sub>1</sub> fingerprints of <sup>32</sup>P-labelled HeLa 28S rRNA were prepared as for 18S rRNA. Fig 4.7 shows an autoradiograph of such a fingerprint. The U<sup>1</sup>graticule is much more complex than on 18S long T<sub>1</sub> separations. In a number of cases spots are very poorly separated from each other. It was therefore, impossible to obtain some oligonucleotides in pure form for base composition and pseudouridine analysis.

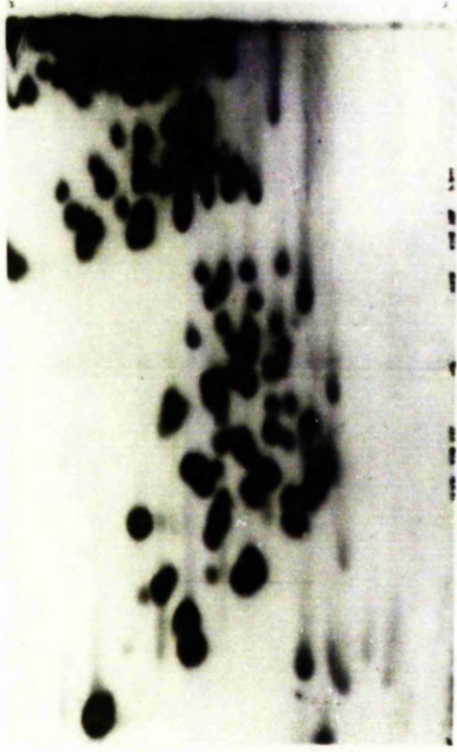
All spots in the one U graticule were subjected to pseudouridine analysis. Low levels of  $\psi$ p were detected in several short sequences which occur a number of times per 28S molecule. Almost all the Up +  $\psi$ p band from spot 104 is  $\psi$ p. The small amount of Up detected was probably a spillover from the adjacent spot 103. A number of the poorly resolved low yield spots above the dashed line in Fig. 4.7b were found to contain a low percentage of  $\psi$ p. The results obtained were, however, extremely variable. It is likely that this complex array of spots includes several  $\psi$ p-containing sequences. Exact localization of the positions of these sequences and quantitation of them is not possible in this system.

Because of the difficulty in obtaining good separation of oligonucleotides a detailed analysis of  $\psi$ p containing sequences in HeLa 28S rRNA was not attempted. However, since  $\psi$ p conversion takes place largely, if not wholly, upon ribosomal precursor RNA, it is

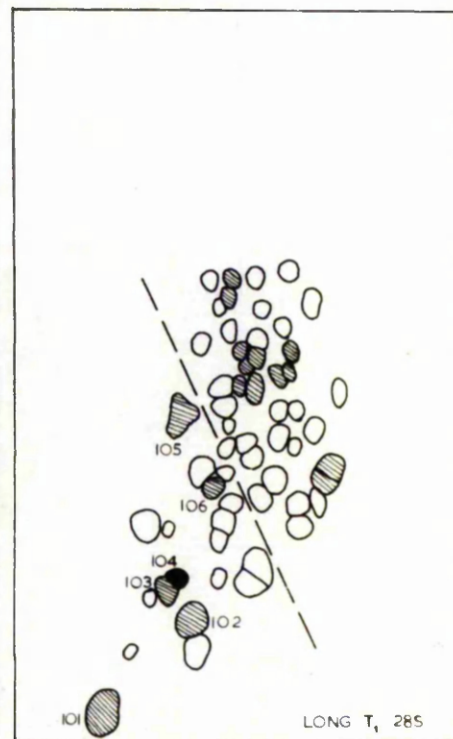
Fig. 4.7

(a) Long  $T_1$  RNase fingerprint of  $^{32}\text{P}$ -labelled HeLa 28S RNA. Separation conditions are described in section 2.4.

(b) Key for "one U" graticule. Products above and to the right of the interrupted line are penta- and larger oligonucleotides, many of which occur once per molecule. Not all the unique products are resolved from each other. Open circles, no pseudouridylate, closed circles, pseudouridylate content suggest one pseudouridylate in a unique sequence. Cross-hatched circles, multiple yield products or incompletely resolved mixtures with some pseudouridylate present.



a.



b.

likely that any general conclusions obtained from analysis of material from 18S RNA would be equally valid for 28S RNA.

#### 4.4. Discussion

In this chapter studies on the location and quantitation of pseudouridine in oligonucleotides of HeLa rRNA have been described. These studies were limited in their scope.

Accurate quantitation of the  $\Psi$ p content of oligonucleotides requires (a) a system of preparation of pure oligonucleotides and (b) a method of complete resolution of Up from  $\Psi$ p. The first requirement is fulfilled partly, but not wholly, by two-dimensional fingerprinting techniques. Some of the spots in the fingerprints shown in Figs. 4.1, 4.6 and 4.7 are clearly separated from all neighbouring spots. However, many of the spots are not completely resolved. This is particularly obvious in the complex pattern obtained in 28S long  $T_1$  fingerprints. Because of incomplete resolution of spots it is not possible to obtain pure samples of some oligonucleotides for subsequent analysis. In the case of 28S spots, cross-contamination resulted in low levels of pseudouridine being detected in several adjacent spots. Exact location of  $\Psi$ p containing sequences was not possible. The problem was less pronounced in 18S fingerprints because of the smaller size of the molecule and smaller number of spots. It was not, however, possible to obtain pure samples of all spots in the fingerprints.

In Chapter 3 a method for the complete resolution of Up from  $\Psi$ p in hydrolysates of whole rRNA was described. This method involves two dimensional chromatography on Whatman 52

paper. The method used in this chapter for separation of uridine and pseudouridine employs the same solvent but chromatography is in one dimension only. Resolution is sufficient to make a reasonable estimate of the  $\Psi$ p content of single oligonucleotides by this method. More accurate quantitation would require two-dimensional separations, but it is doubtful whether the effort involved would have been worthwhile with respect to the two questions posed at the beginning of this chapter.

It is clear from the present results that pseudouridine is present in some "U" containing sequences in HeLa rRNA but is absent from others. Pseudouridine is not spread randomly through the RNA sequence. As in tRNA, modification of Up to  $\Psi$ p appears to occur at specific sites rather than at a low frequency in all Up containing sequences. This conclusion confirms and considerably extends the findings of a previous study on pseudouridine in HeLa rRNA (Amaldi & Attardi, 1968). Among the shorter products which occur in high yield, only approximate estimates of the molar frequencies of the  $\Psi$ p containing sequences were obtained. Results presented here also show that  $\Psi$ p is present in certain long, uniquely occurring sequences. The amount of pseudouridine found in these sequences suggests the complete conversion of a uridine residue to pseudouridine each time the sequence occurs (e.g. spots 159, 207, 305). Thus it appears that pseudouridylation is in many cases an "all-or-none"

modification.

The objectives of sequence analysis of pseudouridine-containing sequences were limited to identification of the shorter sequences. Products of digestion with pancreatic ribonuclease and with snake venom phosphodiesterase were analysed. Attempts to obtain sequence information on the larger pseudouridine-containing sequences by partial digestion with snake venom phosphodiesterase were not persevered with. A major problem with this type of investigation is the large number of stages involved. Accurate location and quantitation of  $\psi$ p in partial digestion products requires a high level of radioactivity. It is difficult (though not impossible), to obtain unique oligonucleotides of sufficiently high activity for this type of analysis. The limited nature of the sequence analysis does not, however, detract from the answers to the major questions on pseudouridine in individual oligonucleotides.



## Chapter 5

### Partial Digestion of 28 S rRNA with $T_1$ Ribonuclease

In Chapter 3 a possible numerical correlation between pseudouridine and 2'-O-methylation was investigated. Chapter 4 described experiments to determine whether the formation of pseudouridine, like 2'-O-methylation, is sequence specific. The studies described in these chapters have given some insight into the presence of modified nucleotides in whole rRNA molecules.

A number of groups have digested 28S rRNA with  $T_1$  ribonuclease under mild conditions in order to prepare large RNA fragments (Gould, 1966, 1967; Wikman et al, 1969; Gould et al, 1966; Pinder et al, 1969; Cox et al., 1973; Godwin et al, 1974). These large  $T_1$  resistant fragments have been shown to be atypical of the 28S RNA molecules from which they are derived, both in base composition and secondary structure (already noted in Section 1.2). No extensive information on the presence of secondary nucleotide modifications in the fragments is available. It seemed possible that  $T_1$  resistant fragments might also be unusual in this respect. This chapter describes the characterization of a large fragment from HeLa rRNA. Particular attention was paid to the presence of pseudouridine residues and 2'-O-methyl groups.

#### 5.1 Purification and Properties of Fragment A

##### 5.1.1 Preparation

The size and number of nuclease resistant fragments

produced by partial digestion of 28S rRNA depends upon the digestion conditions used. Digestion conditions employed in this study were based on those described by Cox et al (1973). Preliminary experiments were carried out by R. Miller as part of an Honours Degree in Biochemistry.

$^{32}\text{P}$  labelled HeLa 28S rRNA was digested with  $\text{T}_1$  RNase with a low Enzyme to Substrate ratio at  $20^\circ\text{C}$  (standard conditions described in Methods). The products of digestion were fractionated on sucrose density gradients and the optical density profile at 260 nm monitored (Fig. 2.6). Radioactivity in each of the fractions was assayed in toluene/triton scintillant. The graphs of radioactivity and optical density follow each other closely.

The fastest sedimenting material on sucrose gradients runs as a peak of optical density partially separated from smaller material. Separation conditions were adjusted to maximize the resolution of this peak. The peak was pooled and designated fragment A. Preparations of fragment A were rerun on sucrose gradients in parallel with unfractionated 28 S partial digests. The purified fragment runs as a single peak at the same position as fragment A in the previously unfractionated digest. The rerun fragment peak is broader than in the initial fractionation but there is no evidence of major heterogeneity.

Digestions of 28S rRNA were also carried out in which

enzyme concentrations 0.1 times and 10 times normal were used. The optical density profiles obtained were apparently identical with those obtained when digestion was carried out under standard conditions. However, a decrease in RNase activity was notable when the same enzyme preparation was used over a period of several months. The separation of fragment A from the smaller material on sucrose gradients became less distinct.

The amount of radioactivity pooled in fragment A was calculated. In four determinations a mean value of 18% of total radioactivity on the gradient was found to be in fragment A. If fragment A represents this percentage of the whole 28 S rRNA molecule, it contains approximately 900 nucleotides.

#### 5.1.2 Base composition

The base composition of fragment A from several preparations was determined. Autoradiographs of electrophoretic separations of T<sub>2</sub> RNase digests show very little material running outside the mononucleotide bands. The nucleotide composition of material running within the mononucleotide bands is shown in Table 5.1. Data show that fragment A preparations contain a very high proportion of cytidine and guanine residues. The conditions of preparation of fragment A were reproduced as exactly as possible on each occasion. Variations in digestion temperature and ribonuclease activity may slightly alter the degree of digestion of RNA. In addition there may be some slight alteration in the exact portion

TABLE 5.1

Nucleotide composition of Fragment A from HeLa cell 28 S rRNA

	%		
Cp	40.4	$\begin{smallmatrix} + \\ - \end{smallmatrix}$	3.4
Ap	9.2	$\begin{smallmatrix} + \\ - \end{smallmatrix}$	1.6
Gp	39.1	$\begin{smallmatrix} + \\ - \end{smallmatrix}$	2.4
Up	11.6	$\begin{smallmatrix} + \\ - \end{smallmatrix}$	2.0

These results are means of 8 determinations using  
T<sub>2</sub> Ribonuclease. Only material running within  
the mononucleotide bands was included.

of the gradient which is pooled as fragment A.

### 5.1.3 T<sub>1</sub> Ribonuclease Fingerprints of Fragment A

Fragment A was completely digested with T<sub>1</sub> RNase and the products fingerprinted (Fig. 5.1). Electrophoresis in the second dimension was for 16 hours. Oligonucleotides in the "no U" graticule are retained on the fingerprint using this length of separation.

The following points may be noted from T<sub>1</sub> fingerprints of fragment A:- (1) There are fewer spots than in T<sub>1</sub> fingerprints of whole 28 S RNA (c.f. 1 U and 2 U graticules of fingerprints in Chapter 4).

(2) Most of the radioactivity is in the short sequences in the "no U" and "one U" graticules. This is consistent with the fact that a large number of G residues is present.

(3) The graticules are heavily biased towards Cp containing oligonucleotides. Ap containing spots are faint and spots containing more than two Ap residues appear to be virtually absent.

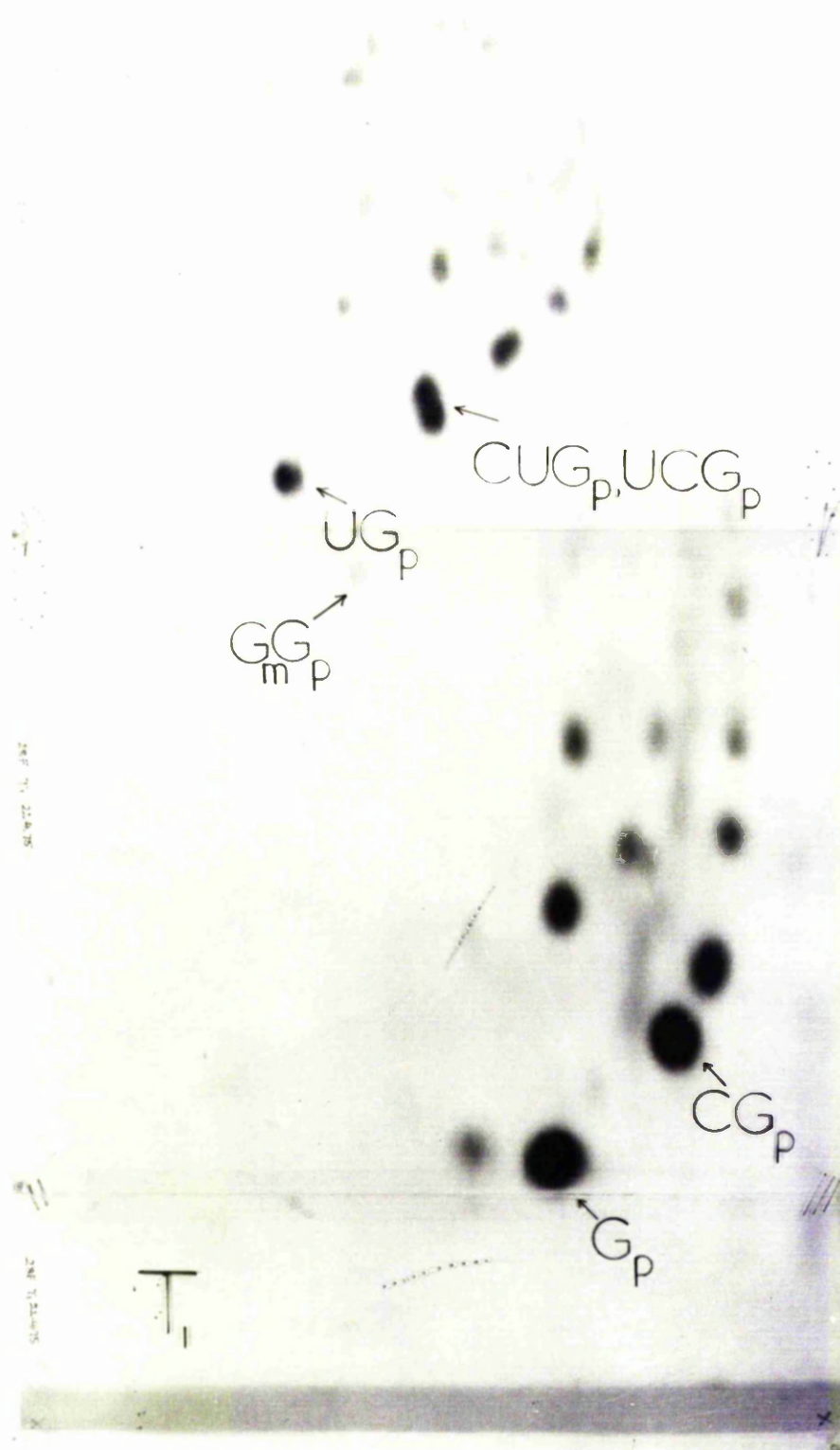
(4) There is very little material which does not migrate in the second dimension of electrophoresis. Many of the spots near the second dimension origin are very faint and may be submolar contaminants.

(5) The only well resolved methylated spot is faint G<sub>m</sub>G<sub>p</sub>. A number of other methylated spots which would be resolved

Fig. 5.1

Standard T<sub>1</sub> RNase fingerprint of HeLa fragment

A. Conditions of digestion and electrophoresis are  
described in section 2.4.



in  $^{32}\text{P}$  fingerprints are absent. These observations are consistent with the fact (Table 5.1) that the fragment is highly rich in Gp and Cp residues and suggests that few methyl groups are present (investigated in more detail later). The preparation appears to be reasonably pure although some faint contaminating spots are detectable.

#### 5.1.4 Pancreatic Ribonuclease Fingerprints of Fragment A

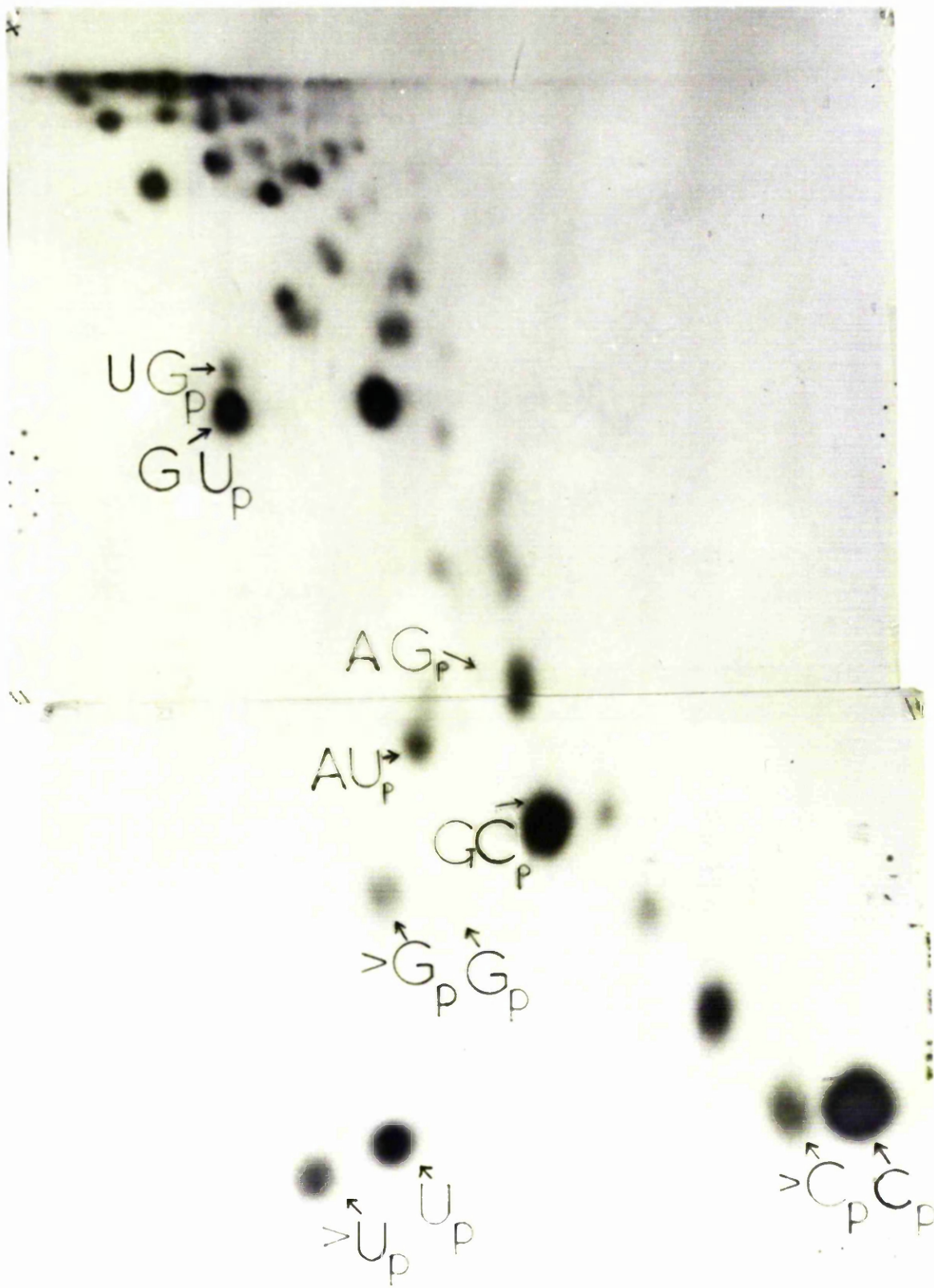
Pancreatic ribonuclease cleaves RNA at Cp and Up residues only. Fig. 5.2. shows the fingerprint of a pancreatic RNase digest of HeLa fragment A. In addition to the expected oligonucleotides terminated by Cp and Up, spots due to Gp and cyclic Gp are present as well as traces of AGp and UGp. These can only be due to digestion of 28 S RNA at certain Gp residues in the original partial  $T_1$  RNase digests:- An RNA species with a high degree of secondary structure can undergo digestion at a few points within its sequence without loss of apparent integrity. The various oligonucleotides present are held together by base-pairing. Subsequent digestion with pancreatic RNase releases oligonucleotides terminated by Gp from positions adjacent to internal nicks. The presence of Gp, UGp and AGp in pancreatic fingerprints suggests that the fragment contains a number of these internal nicks.

Table 5.2 shows the relative yields of spots in Pancreatic RNase fingerprints of Fragment A. Yields were expressed as a percentage of that of the strongest spot Cp (plus cyclic Cp). It



Fig. 5.2

Pancreatic RNase fingerprint of  $^{32}\text{P}$ -  
labelled HeLa fragment A. Conditions of digestion  
and separation are described in section 2.4.



**P**<sub>anc</sub>

TABLE 5.2.

Relative Yield of Spots in Pancreatic RNase Fingerprints  
of Fragment A.

A figure of 100 was arbitrarily set for the yield of the strongest spot  
Cp (plus cyclic Cp).

<u>Spot</u>	<u>Yield</u>
Cp plus Cyclic Cp	100
ACp	5.4
A <sub>2</sub> Cp	0.4
A <sub>3</sub> Cp	0.2
Up plus cyclic Up	21.7
AUp	0.5
A <sub>2</sub> Up	0.2
Gp plus cyclic Gp	2.3
GCp	23
GUp	7.0
UGp	1.5
AGp	0.1

is not possible to estimate the exact number of internal nicks caused by  $T_1$  since the molarity of none of the spots is known. However, the combined yield of Gp, AGp and UGp is only 4% of that of Cp.

#### 5.1.5 Polyacrylamide Gel Electrophoresis of Fragment A

Fig. 5.3 shows an electrophoretic separation of  $^{32}\text{P}$ -labelled fragment A on a neutral 6% polyacrylamide gel. There are five prominent bands running near the top of the gel of which band 3 is the most radioactive. A number of weakly labelled bands are visible in the lower portion of the gel.

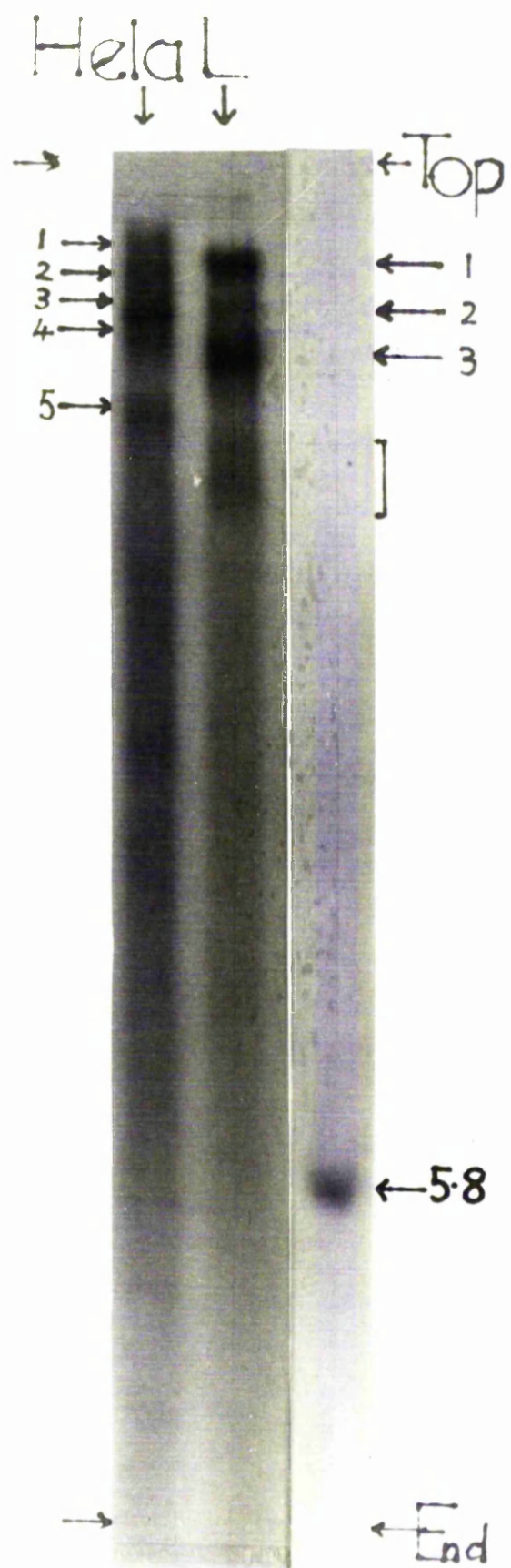
The relative strength of the five major bands was found to vary somewhat between different fragment A preparations. In particular the relative strength of band 1 varied markedly. This probably reflects a variation in the portion of the gradient pooled as fragment A. Band 1 is the largest species present and may run slightly ahead of the main peak of optical density on gradients. The amount of band 1 material collected may depend on exactly where collection of fragment A is commenced for each preparation.

18S and 5.8S RNA's were run on a 6% gel in parallel with fragment A. 18 S hardly migrated into the gel whilst 5.8S migrated much faster than any of the major bands in the fragment preparation. In the absence of suitable size markers it is not possible to calculate exact sizes of the species in each of the main bands. All the major fragments are much larger than 5.8S RNA (160 nucleotides) but

Fig. 5.3

Electrophoresis of HeLa and L-cell fragment A preparations on neutral 6% polyacrylamide gels.

The positions of the major bands are marked with arrows - HeLa on the left, L-cell on the right. HeLa 5.8S RNA was run alongside the fragment preparations and its position is also marked.



none is as large as 18 S RNA (2000 nucleotides). The estimate of 900 nucleotides length obtained from radioactivity measurements probably represents a reasonable estimate of the mean size of the species present.

## 5.2 Secondary structure of Fragment A

In section 5.1 it was shown that fragment A has a high G + C content. Although internal nicks are present in its sequence the integrity of the fragment is maintained. This suggests that fragment A has considerable secondary structure based mainly on the formation of G-C base pairs. In subsequent sections experiments are described in which the secondary structure of fragment A was further investigated

### 5.2.1 Denaturing Gels

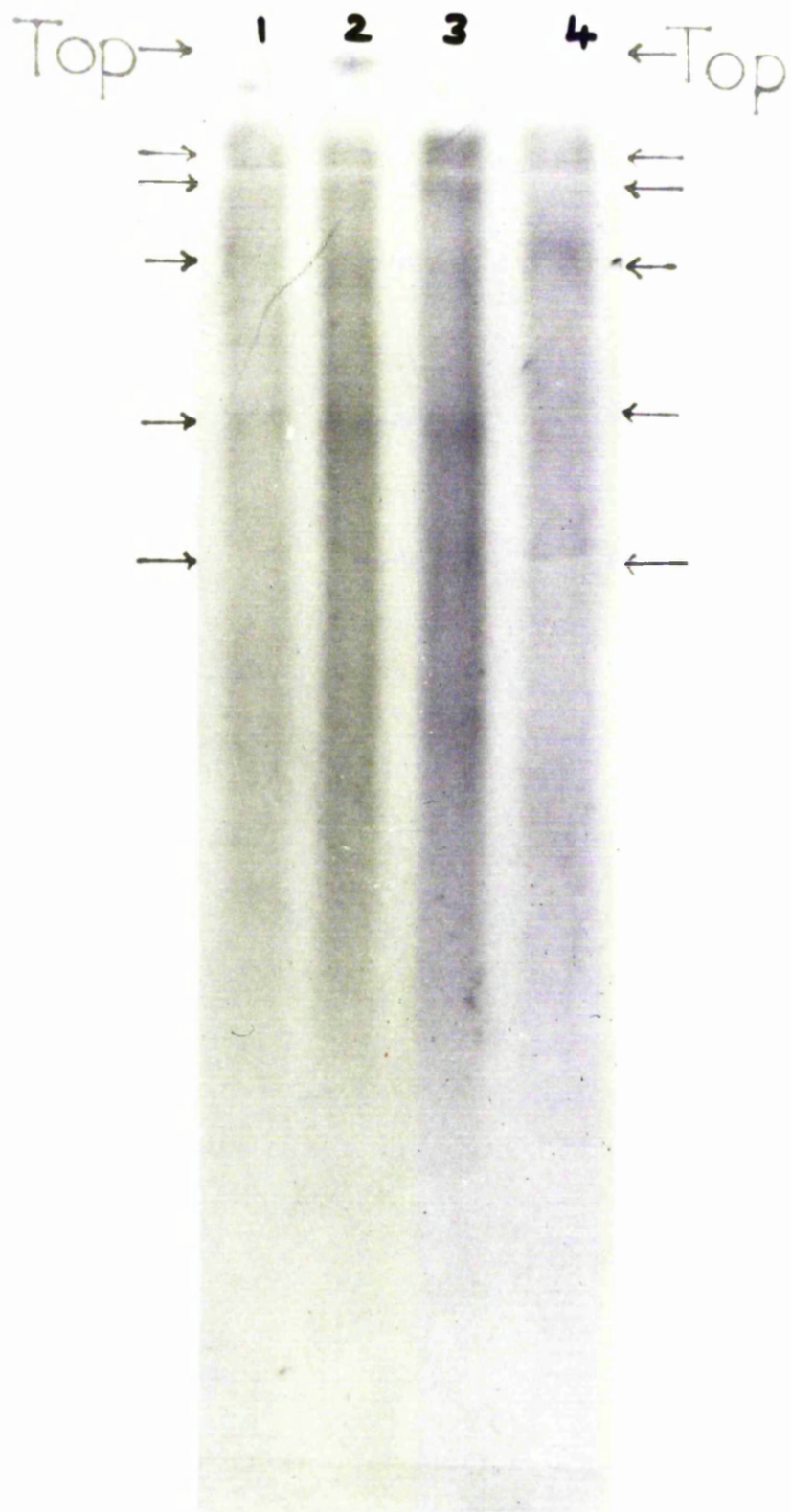
Bands from neutral polyacrylamide gels (Section 5.1) were cut from the original gel slab and run on a 9% polyacrylamide, 6M urea denaturing gel. Fig. 5.4 shows an autoradiograph from this type of gel. The background radiation is quite high on this separation and visualization of faint bands is difficult. Comparison of the band patterns in each slot shows a high degree of similarity. The intensity of the major bands varies somewhat between slots but there are several bands which are common to all.

Denaturing gels confirm the presence of a number of internal nicks in the fragment A sequence. The similarity of pattern of all the main bands from neutral gels suggests that they may have part of their sequence in common. Fragment A

Fig. 5.4

Electrophoresis of HeLa fragment A bands from neutral 6% polyacrylamide gels on 6M urea, 9% polyacrylamide gels. The positions of the major bands are marked with arrows.





preparations may be a collection of fragments which differ somewhat in length but which are all derived from the same part of the 28 S rRNA molecule.

### 5.2.2 Bisulphite Modification

Sodium bisulphite reacts with non-hydrogen bonded pyrimidine residues. In particular, cytidine forms an adduct (5 - 6 dihydrocytidine -6 - sulphonate) which is non-reversibly converted to uridine at slightly elevated pH. This reaction has been used to identify non-hydrogen-bonded cytidine residues in tRNA (Goddard & Schulman, 1972). More recently Goddard & Maden (1976) have used the reaction as a conformational probe for methylated sequences in HeLa 28 S rRNA.

Bisulphite modification was used here as a probe for secondary structure in fragment A. The fragment is rich in Gp and Cp residues when compared with whole 28S RNA. If it is also relatively rich in G-C base pairs its reactivity to sodium bisulphite should be low when compared with that of 28 S RNA.

HeLa fragment A and major 28 S rRNA were reacted with 3M bisulphite, pH 6.0 at 25°C. The time course of the reaction was followed over several days and base compositions of samples withdrawn at various times were determined. (Table 5.3). From the results obtained a graph of log (remaining C) vs. time was plotted (Fig. 5.5).

The results obtained for major 28 S rRNA are in reasonable

TABLE 5.3

(a) 28 S

Time/hours	%				
	Ap	Gp	Up + Up	Cp	% Remaining C
0	16.3	35.6	17.5	30.6	100
24	16.4	35.8	23.4	24.5	79.9
48	16.2	35.7	26.2	21.9	71.3
96	15.5	35.7	31.3	17.6	57.4

(b) Fragment A

Time/hours	%				
	Ap	Gp	Up + Up	Cp	% Remaining C
0	9.7	38.6	12.5	39.2	100
24	8.8	39.1	19.0	33.1	84.4
48	8.9	39.3	20.7	31.1	79.3
96	9.0	38.2	25.6	27.2	69.3

Change in Nucleotide Composition After Reaction with Bisulphite

RNA was reacted for the periods indicated, subjected to complete hydrolysis with  $T_2$  RNase and the products separated by electrophoresis on Whatman 52 paper at pH 3.5. Only radioactivity in the four major bands was determined. The values shown are means of three determinations.

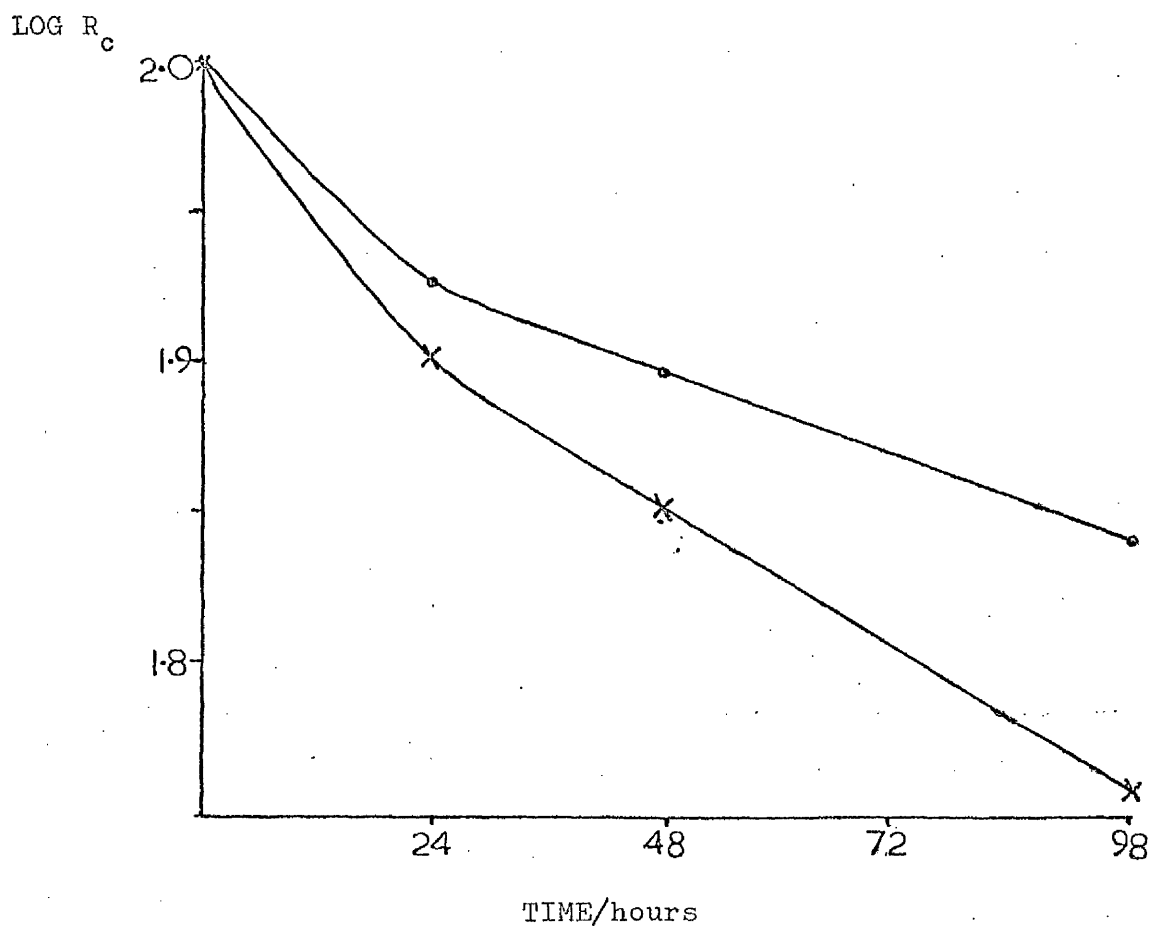


Fig. 5.5 Graph of  $\log(\% \text{ Remaining } C_p)$  vs. time for modification of Hela fragment A and 28S rRNA with 3M bisulphite, pH 6.0 at 25°C. Aliquots were withdrawn at various times and their base composition analysed. The  $\% C_p$  at any given time was expressed as a percentage of the  $\% C_p$  in the sample withdrawn at  $t=0$ .

agreement with those of Goddard and Maden (1976). Reaction is fastest during the first 24 hours. Even after 96 hours no plateau in the reaction has been reached. The reactivity of fragment A to sodium bisulphite is markedly less than that of the major 28 S molecule from which it is derived. Conversion of 30% of Cp residues takes about twice as long (ca.96 hours) in fragment A as in 28S RNA (ca.48 hours). This suggests that a greater proportion of Cp residues in fragment A is involved in hydrogen-bonding interactions. The reactivity of 28S RNA to bisulphite is a combination of the reactivity of the fragment A part of the molecule and that of other parts of the molecule. It seems therefore, that the parts of 28S RNA not present in fragment A must be relatively more reactive to bisulphite and contain relatively few G-C pairs.

### 5.2.3 Electron Microscopy of HeLa Fragment A.

Electron microscopy has given considerable insight into the secondary structure of rRNA (Discussed in Chapter 1). The major structural features of HeLa 28S rRNA have been determined by heteroduplex mapping (Wellauer & Dawid, 1974). Fragment A is derived from HeLa 28S rRNA and is thought to contain a considerable degree of secondary structure. It therefore seemed possible that electron microscopy of fragment A might reveal interesting structural features.

Fragment A was spread for electron microscopy in 80% formamide (see methods). Electron micrographs (Fig. 5.6) show a collection of branched molecules in a variety of

Fig. 5.6

Electron micrograph of HeLa fragment  
A (X 150,000). Conditions of spreading are  
described in section 2.6.



0.25 $\mu$



Number  
counted

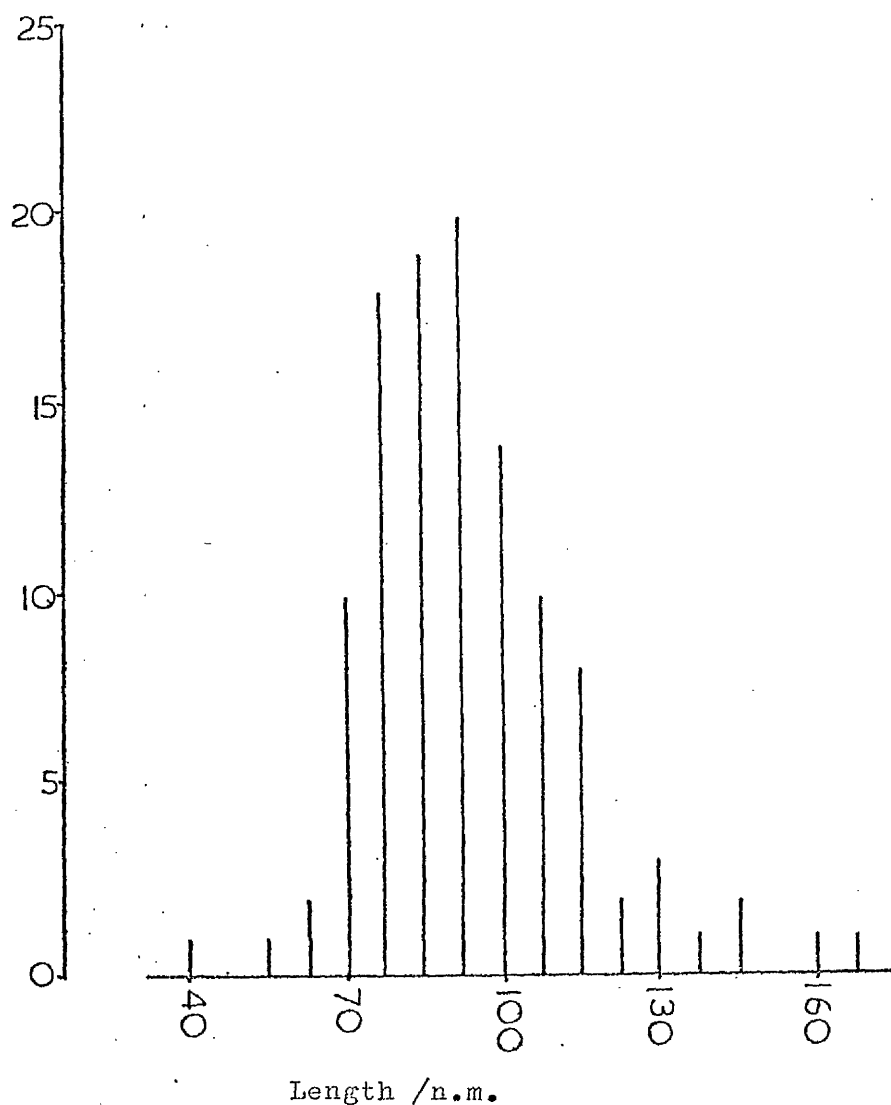


Fig. 5.7 Histogram of lengths of fragment A molecules as seen in electron micrographs. The contour lengths of 114 molecules in four different micrographs were measured. The actual length of molecules was calculated from the magnification of the electron microscope.



configurations. The most common species appears to consist of one short branch and two longer branches of roughly equal length. A few short unbranched molecules are seen which may be single branches cleaved from larger species.

Molecules in several electron micrographs were enlarged with a profile projector and their contour lengths measured. Molecular size was calculated approximately from the nominal magnification of the microscope. No internal length standard was included in the micrographs. Fig. 5.7. shows a histogram of the contour lengths measured. Most of the molecules have contour lengths of between 70 nm. and 120 nm. This corresponds to a mass of approximately  $0.15 \times 10^6$  to  $0.26 \times 10^6$  Daltons (450-800 nucleotides) if the molecules are completely bihelical.

The high G + C content of fragment A and its low reactivity to bisulphite suggest a high degree of base-pairing. Although base-pairing is extensive it may be imperfect. As described above, pancreatic RNase fingerprints and denaturing gels have shown that a number of internal nicks are present in the fragments. The denaturing solvent used for electron microscopy may cause some breakdown of secondary structure around these nicks. It is, therefore, probable that the molecules observed in electron micrographs are not perfectly bihelical. Estimates of the molecular weight of the fragment may be too large because of this.

### 5.3. Modified Nucleotides in Fragment A

#### 5.3.1 Methylation

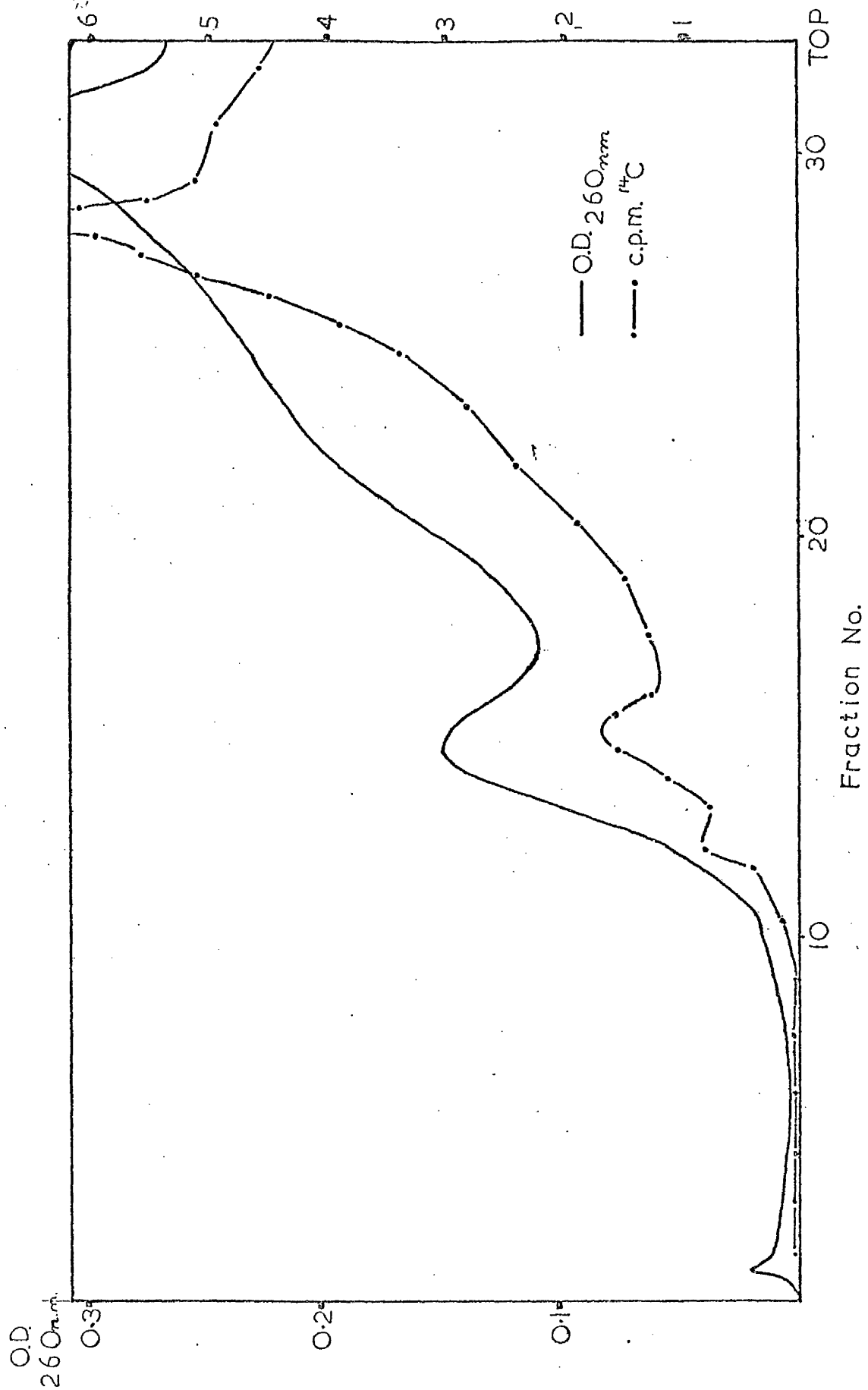
Fingerprints of pancreatic RNase and  $T_1$  RNase digests of fragment A have suggested that only a few methyl groups are present. Digestion with a combination of  $T_1$  and pancreatic RNases is particularly useful when studying methylation. Fingerprints of this type are fairly simple. Many characteristic methylated sequences are well resolved from unmethylated sequences in  $^{32}\text{P}$  fingerprints due to the presence of enzyme-resistant methylated linkages. Methylation in fragment A has been studied by use of  $T_1$  plus pancreatic fingerprints of both  $^{14}\text{C}$  and  $^{32}\text{P}$  labelled material.

Partial digestion of  $^{14}\text{C}$ -methyl labelled HeLa 28 S rRNA was carried out as usual and the products fractionated on sucrose density gradients. Fig. 5.8 shows the distribution of radioactivity across the gradient. The graph of  $^{14}\text{C}$  counts does not closely follow the optical density profile as was the case in  $^{32}\text{P}$  digests. Most of the radioactivity is in material sedimenting more slowly than fragment A. There is a small peak of activity corresponding to the major part of fragment A. In addition, there is a small faster running peak which corresponds to the leading edge of the fragment optical density peak. In five preparations a mean value of  $1.9 \pm 1.2\%$  of total radioactivity was found in this small peak whilst  $4.7 \pm 1.6\%$  was pooled in Fragment A.

$^{14}\text{C}$  labelled fragment A was completely digested with

Fig. 5.8

Distribution of radioactivity in partial digests of  $^{14}\text{C}$ -methyl labelled HeLa 28S RNA. RNA was digested under standard conditions and products separated by sucrose gradient centrifugation. The optical density of the fractions was monitored and aliquots were assayed for  $^{14}\text{C}$ -radioactivity in toluene/triton scintillant.



$T_1$  and pancreatic ribonucleases and fingerprinted. The autoradiograph of the separation shows only one clear spot (Fig. 5.9a). This spot was excised, digested with RNase  $T_2$  and subjected to electrophoresis at pH 3.5 (Fig. 5.9b). A single spot was seen whose mobility was intermediate between Gp and Up marker mononucleotides suggesting that Gm-Gp is present (Maden & Salim, 1974). In a subsequent  $T_1$  plus pancreatic fingerprint of  $^{14}\text{C}$ -labelled fragment A a second spot was observed running close to GmG. This faint spot was probably due to Um-Gp and/or Gm-Up.

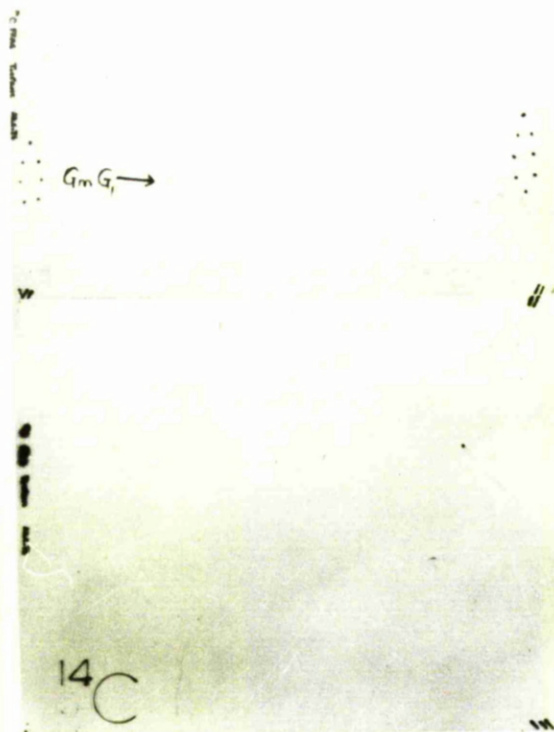
Fig. 5.10 shows  $T_1$  plus pancreatic fingerprints of  $^{32}\text{P}$  labelled fragment A and of whole 28 S rRNA. The pattern from 28 S RNA shows a number of small spots due to methylated oligonucleotides (characterized by Maden & Salim, 1974).  $T_1$  plus pancreatic fingerprints of fragment A are much simpler than for 28 S RNA. Cp and Gp are obviously present in much larger amounts than Up. The only methylated spots which can be seen are GmG and faint UmG. It is not possible to determine the presence or absence of those few methylated products which run coincidentally with non-methylated oligonucleotides in this system.

The exact number of methylations present in fragment A cannot be accurately determined from these results. It is however clear that the level of methylation is low and that many characteristic 28 S methylations are absent. It appears that the proportion of total  $^{14}\text{C}$ -methyl counts in the fragment A portion

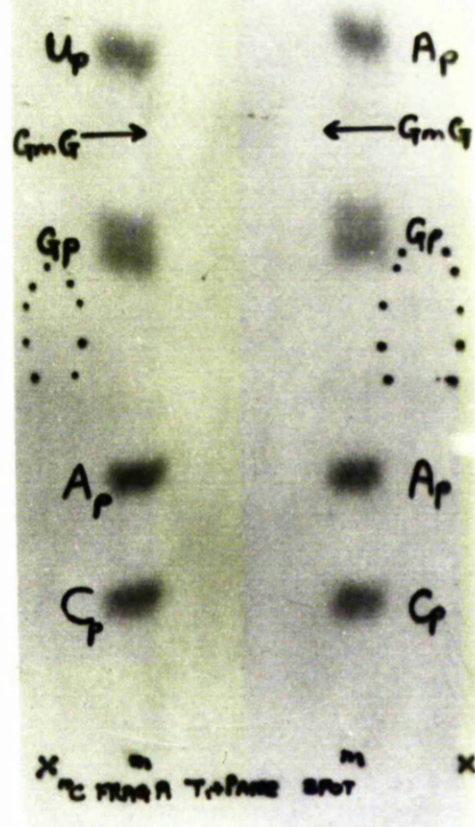
Fig. 5.9

(a)  $T_1$  plus pancreatic RNase fingerprints of  $^{14}\text{C}$ -labelled HeLa fragment A.

(b) pH 3.5 electrophoretic separation of alkali digested material from spot marked GmG in (a).



a.



b.

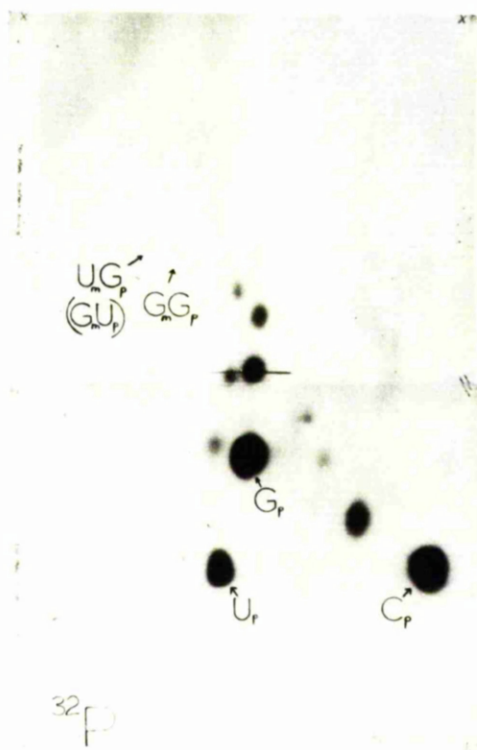
Fig. 5.10

$T_1$  plus pancreatic RNase fingerprints of

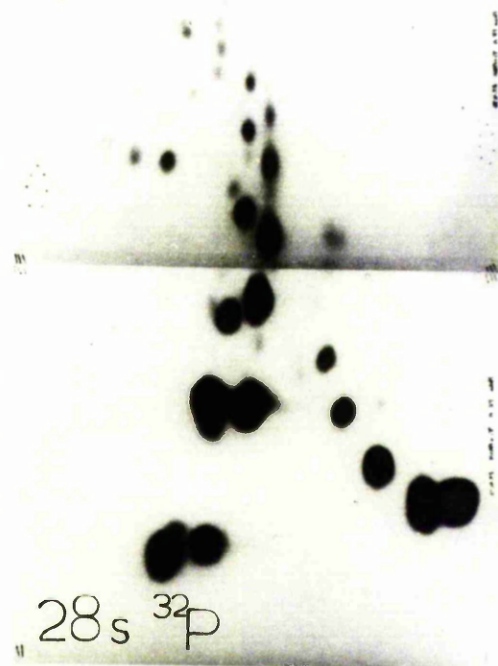
(a)  $^{32}\text{P}$ -labelled HeLa fragment A.

(b)  $^{32}\text{P}$ -labelled HeLa 28 S rRNA.





a.



b.

of sucrose gradients is much less than the proportion of  $^{32}\text{P}$  counts in the corresponding fractions. This suggests that methylation is significantly less frequent in the fragment A portion of the 28 S sequence than in the sequence as a whole.

The spots in  $T_1$  plus pancreatic fingerprints of  $^{32}\text{P}$  fragment A were assayed for radioactivity. The counts per minute in two low yield spots ( $A_2U$  and  $A_2C$ ) were virtually identical. These sequences were assumed to occur once each in the fragment sequence. All other oligonucleotides were quantitated relative to these sequences by use of the following relationship :-

$$A_2U - x \quad \text{c. p. m.}$$

$$A_2C - y \quad \text{c. p. m.}$$

$$\text{Therefore c. p. m. per mole of nucleotide} = \frac{x + y}{6}$$

suggested molar yields are shown in table 5.4. The methylated dinucleotide Gm-Gp is apparently almost unimolar. Um-Gp/Gm-Up occurs in consistently less than unimolar amounts. The presence, demonstrated by pancreatic RNase fingerprints and gels of internal nicks shows that some regions of fragment A are somewhat exposed to nuclease attack. Slight variations in molecular configuration in such regions might make loss of sequences adjacent to nicks possible in some cases. It seems likely that the sequence Um-Gp/Gm-UP occurs in a position which sometimes makes it liable to removal in the initial  $T_1$

TABLE 5.4Analysis of Fragment A T<sub>1</sub> + Panc Counts

<u>Spot</u>	<u>Suggested Molar Yield</u>	
	<u>HeLa</u>	<u>L-cell</u>
Up	80.5	93.0
AUp	4.3	4.6
A <sub>2</sub> Up	1.0	1.2
Cp	426.0	409.2
A Cp	15.5	16.3
A <sub>2</sub> Cp	1.0	0.8
A <sub>3</sub> Cp	0.5	0.4
Gp	378.2	388.6
AGp	19.7	18.8
A <sub>2</sub> Gp	3.8	3.8
Gm-Gp	0.8	0.7
Um-Gp/Gm-Up	0.3	0.2
Total	<u>931.6</u>	<u>937.6</u>

These results are means from three fingerprints for  
HeLa and two for L-cells.

partial digestion.

Addition of the suggested molar yields of oligonucleotides in Table 5.4 suggests that fragment A contains some 930 nucleotides. This figure is larger than other estimates of the size of fragment A but not greatly so (see Table 5.6). It appears that the assumptions used in calculation of molar yields are, at least approximately, correct.

### 5.3.2 Pseudouridine

It was of interest to determine whether the numerical correlation between pseudouridine and 2'-O-methyl groups in HeLa rRNA was valid for a particular part of the sequence. Fragment A, which is derived from HeLa 28S rRNA, evidently contains very few methyl groups. If pseudouridylation and methylation are in some way associated, the number of  $\psi$ p residues in the fragment should also be low. This possibility was investigated by assaying the pseudouridine content of fragment A, as well as other fractions from separations of 28 S RNA partially digested with  $T_1$  RNase.

$^{32}\text{P}$  labelled HeLa 28 S rRNA was partially digested with  $T_1$  RNase and fractionated as previously. Fractions A to E were pooled as shown in Fig. 5.11. These fractions were then assayed for pseudouridine content using the two-dimensional chromatography system as described for whole rRNA.

Table 5.5. shows the pseudouridine content of

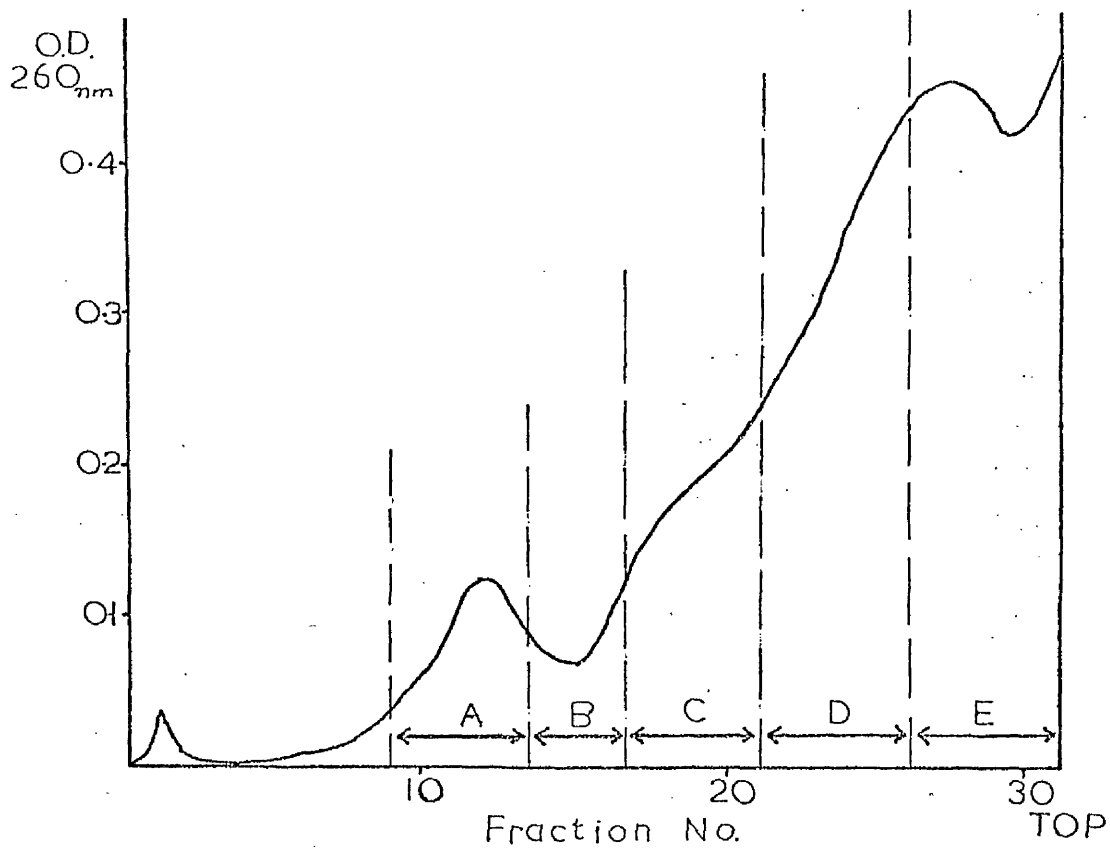


Fig 5.11 Fractionation of material from partial  $T_1$  RNase digests of HeLa 28S rRNA for pseudouridine determination. Fractions A to E were collected as shown and analysed for pseudouridine content.

TABLE 5.5

Pseudouridine Content of Partial Digest Gradient Fractions

Fraction	%	$\frac{\psi p}{U_p + \psi p}$
A		2.4
B		3.8
C		5.4
D		7.9
E		10.2
Whole 28 S RNA		7.3

Gradient fractions were pooled as shown in Fig. 5.11.

Values shown are means of two determinations.

fractions A to E. The level of pseudouridine increases markedly on going from larger to smaller RNA fragments. Fragment A has a low  $\psi$ p content both in comparison with other partial digest fractions and with whole 28S rRNA (in which 7.3% of Up +  $\psi$ p is  $\psi$ p). The total number of pseudouridine residues in fragment A was calculated using its known base composition and the various size estimates. This suggests that between one and three pseudouridines are present in fragment A.

#### 5.4 Partial Digestion of Other rRNA's

Digestion of HeLa 18 S rRNA and fractionation of the products was carried out as for HeLa 28 S rRNA. The optical density profile of the products shows a heterogeneous spread of material. No distinct nuclease resistant species is detectable.

The 28 S rRNA's of L-cells, *Xenopus laevis* and *Dictyostelium discoideum* were digested with  $T_1$  RNase under standard partial digestion conditions. Fig. 5.12 shows the optical density profiles obtained after fractionation of the digestion products on sucrose density gradients.

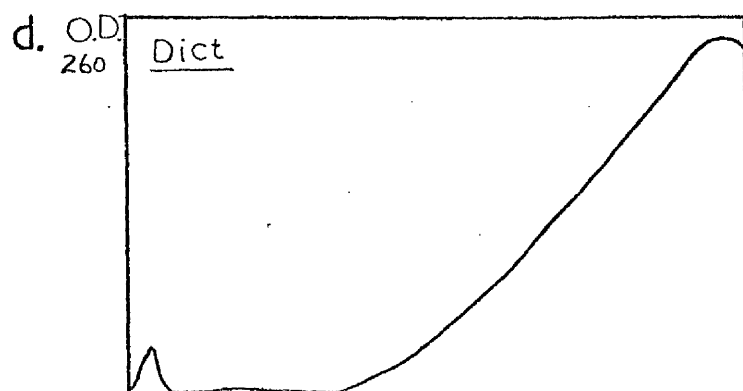
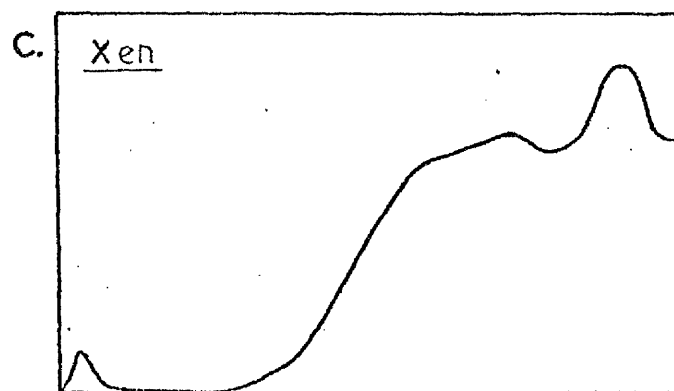
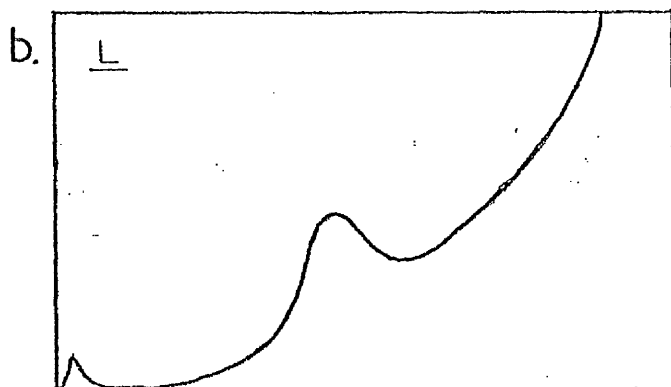
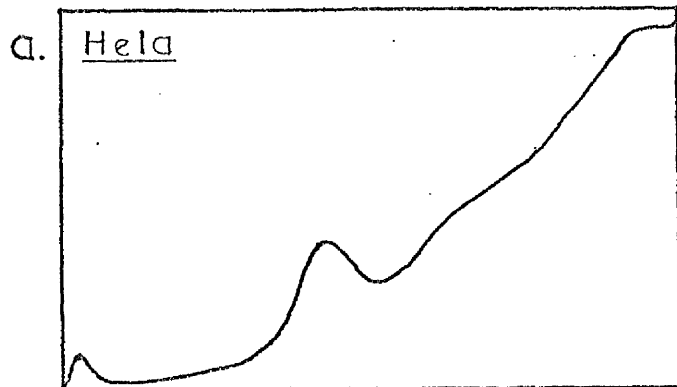
The profile for L-cell RNA is indistinguishable from that of HeLa 28 S rRNA digests. There is a single peak, partially resolved from smaller material. The position of this peak corresponds closely to that of fragment A in a digest of HeLa 28 S RNA run in parallel.

Fig. 5.12

O. D. profiles of the products of  $T_1$  digestion of 28S RNA from (a) HeLa cells; (b) L-cells; (c) *Xenopus laevis*; (d) *Dictyostelium discoidium*.

Digestion was carried out, in each case, using the standard conditions for partial  $T_1$  RNase digestion described for HeLa 28S RNA. The products of digestion were separated on 10 — 25% sucrose-LETS gradients.





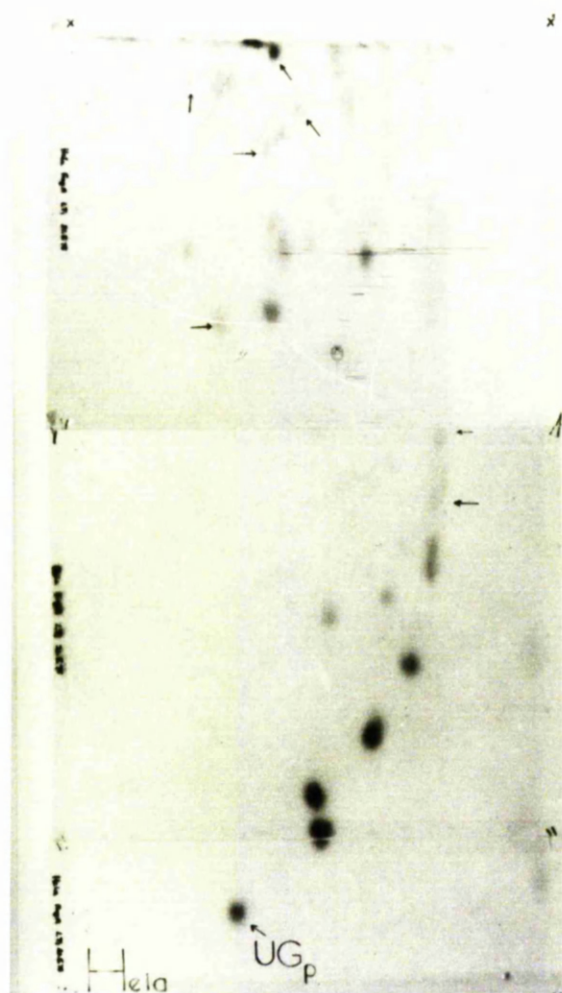
TOP

Fig. 5.13

Long  $T_1$  RNase fingerprints of fragment A from  
(a) HeLa cells and (b) L-cells . Conditions of digestion  
and electrophoresis are described in section 2.4.

The arrows indicate spots which are unique to

(a) the HeLa fingerprint and (b) the L-cell fingerprint.



a.



b.

The optical density profile of *Xenopus* partial digests suggests that the digestion products are not completely heterogenous. There is not, however, any major species resolved from the background. It appears that *Dictyostelium* rRNA is digested to small fragments under the conditions used.

#### Comparison of Fragment A from HeLa and L-cells

L-cell fragment A was run on a 6% polyacrylamide slab gel alongside HeLa fragment A (Fig. 5.2). The L-cell preparation shows three main bands of radioactivity. The size range of the bands from the two preparations is similar. However, the exact distribution of bands differs between the two cell lines.

T<sub>1</sub> plus pancreatic ribonuclease fingerprints of <sup>32</sup>P-labelled L-cell fragment A are visually indistinguishable from those of HeLa fragment A. All the same spots are present in approximately the same yield. The only visible methylated spots are GmG and faint UmG. Suggested molar yields of spots in L-cell fingerprints were calculated as for HeLa (Fig. 5.4). There is some slight differences of distribution of mononucleotides in the two fingerprints but most of the molar yields are very similar.

Long T<sub>1</sub> ribonuclease fingerprints of fragment A from both HeLa and L-cells were prepared. The oligonucleotide patterns of both species are considerably less complex than for whole 28S rRNA (Maden & Salim, 1974). (Fig. 5.13). As expected from base composition

data Cp containing oligonucleotides are obviously present in higher yields than those containing Ap. Although there is considerable overall similarity between the two fingerprints there are a number of spots which occur in one species only (arrowed in Fig. 5.13). There is no evidence of the presence of methylated oligonucleotides in either fingerprint although no detailed analysis of products was undertaken.

In summary, it appears that partial digestion of 28 S rRNA from HeLa and L-cells produces nuclease resistant fragments with a considerable degree of similarity between the species. The fragments produced from the two species display indistinguishable mobility on sucrose density gradients. Polyacrylamide gels show that the preparations are similar but not identical. Fingerprinting studies confirm that fragments from the two species which have nucleotide compositions which are generally similar but which differ in detail.

### 5.5. Discussion

Fragment A preparations appear to consist of a collection of several large fragments which differ somewhat in size. The similarity in the pattern produced by each of these species on denaturing gels suggest that they have some part of their sequence in common. It is possible that all the major species in fragment A preparations are derived from the same part of the 28 S molecule and that they differ in size because of variation in the extent of  $T_1$  digestion at their ends.

Studies of pancreatic RNase fingerprints and denaturing gels show that partial  $T_1$  digestion causes some internal nicks within fragment A molecules. Fragments are largely held together in spite of these nicks, presumably because of secondary-structural interactions. Any disruption of these interactions may cause some part of the fragment to be lost. Limited disruption of base pairing may be one reason for size variation within fragment A preparations.

No conclusive determination of size of the species in fragment A preparations has been made. However, estimates obtained from electron microscopy,  $T_1$  plus pancreatic RNase fingerprints, gel electrophoresis and radioactivity distribution are reasonably consistent (Table 5.6). Combination of these estimates suggests a size range of 0.15 - 0.3 Daltons (450-900 nucleotides). This is within the size range of large fragments produced by mild  $T_1$  digestion of 28S rRNA by two other groups (Wikman et al, 1969; Cox et al, 1973).

Fragment A contains a much higher proportion of Gp and Cp residues (about 80%) than the 28 S rRNA from which it is derived (ca. 67% G+C). Reactivity studies with sodium bisulphite suggest a high proportion of Cp (and Gp) residues in fragment A are involved in base-pairing. The conclusion of extensive base-pairing is supported by electron microscopy. Micrographs show that fragment A molecules are largely double-

TABLE 5.6.

Size Estimates for Fragment A

<u>Experiment</u>	<u>Approximate Number of Nucleotides</u>
<sup>32</sup> P-radioactivity distribution	900
Neutral PAGE	mean 900
Electron Microscopy	450 - 800 .
T <sub>1</sub> plus pancreatic RNase fingerprints	930

The derivation of the various size estimates  
is described in the text.

stranded. It is clear that many G-C base-pairs are present which are stable in the denaturing spreading conditions used. Godwin et al (1974) have carried out electron microscopy on preparations of large  $T_1$  resistant fragments from rabbit reticulocyte 28 S rRNA. The visual appearance of the fragments in electron micrographs was similar to that of the HeLa rRNA fragments described here. Contour lengths measurements suggest that the range of molecular size of fragments in the two studies is approximately the same.

Electron microscopic studies have revealed the presence of atypical regions at defined points within the 28 S rRNA sequence (Reviewed in Chapter 1). These regions are believed to have a high G + C content and considerable bihelical character. The properties of fragment A strongly suggest that it is derived from one of these regions of the 28S molecule. Two major secondary-structural features are revealed in electron micrographs of 28 S rRNA (Fig. 1.1). Godwin et al (1974) suggested the reticulocyte RNA fragment which they were investigating was derived from the complex arrangement of loops at the 5' end of the 28 S sequence. It does, however, seem equally likely that G - C rich fragments arise from the "rabbit's ears" - like feature situated near the mid-point of the 28 S molecule. It is perhaps possible that preparations of  $T_1$  resistant fragments contain species of similar size, from both the major secondary-structural features of 28S rRNA.



Results from two earlier studies (Gould et al, 1966; Pinder et al, 1969) suggested that the size of nuclease resistant G - C rich fragments from 28 S rRNA's increases with the evolutionary complexity. The results of digestion of four eukaryotic rRNA's described here are in agreement with this conclusion. 28 S rRNA from the two higher vertebrates studied gives rise to well defined large fragments. rRNA from the lower vertebrate, *Xenopus* apparently yields only smaller nuclease resistant fragments. *Dictyostelium* rRNA apparently does not produce any resistant fragments of detectable size when digested under the same conditions. The increasing tendency to form resistant fragments follows the trend of increasing G + C content of r RNA in evolution. Secondary-structural loops which are seen in electron micrographs of rRNA also increase in size with increasing evolutionary complexity (Wellauer & Dawid, 1973, 1974; Wellauer et al, 1974). The size of the major loops decreases slightly between HeLa and L-cells and more sharply between L-cells and *Xenopus*. The A-U rich rRNA of *Dictyostelium* does not form any resistant loops. This correlation between the size of T<sub>1</sub> resistant fragments and the size of secondary structural loops reinforces the suggestion that large G-C rich fragments from 28 S rRNA are derived from the major features seen under the electron microscope.

The relative distribution of <sup>14</sup>C-methyl and <sup>32</sup>P radioactivity in partial digest gradients suggests that fragment A

contains less than half the number of methyl groups expected if methylation were evenly distributed through the 28 S molecule. Both  $^{14}\text{C}$  and  $^{32}\text{P}$  fingerprints suggest that almost all of the methylated oligonucleotides characteristic of 28S rRNA are absent from fragment A. Only one or two spots corresponding to methylated sequences are seen. The possibility that the methylated sequences apparently present in Fragment A are due to contamination by material from the rest of the 28 S sequence cannot, however, be completely ruled out. Gm-Gp occurs eight times in whole 28 S RNA (Maden & Salim, 1974). A 1:8 contamination of fragment A by the rest of the 28 S sequence would make it appear that Gm-Gp occurs at a molar frequency of one. Contamination by other methylated sequences which occur only once or twice in the 28 S sequence would probably not be detected. The presence of faint spots in  $T_1$  fingerprints of  $^{32}\text{P}$ -labelled fragment A (Fig. 5.1) suggests that some low level of contamination may be present.

Pseudouridine is heterogeneously distributed through the fractions of the partial digest gradients but the percentage of  $\psi\text{p}$  found in fragment A is only one-third of that in whole rRNA. Fragment A contains a maximum of three pseudouridines whereas 28 S RNA contains approximately sixty. Therefore, fragment A appears to be unlike whole 28 S rRNA in its degree of secondary modification as well as in its primary and secondary structure. Both of the most common types of secondary modification occur

at a frequency much lower than that expected if modifications were distributed evenly through the 28 S molecule. Estimates obtained here suggest that the two types of modification occur at approximately equal frequency in fragment A. This appears to be consistent with the numerical correlation between numbers of methyl groups and pseudouridines suggested for whole rRNA.

## Chapter 6

### Discussion

Particular aspects of the results have been discussed in the relevant chapters. In this chapter data from other investigations will be noted and their possible relevance to results in this study discussed. In addition, reference will be made to other experimental systems which may be of use in the investigation of secondary modification of ribosomal RNA.

#### 6.1. Previous Studies on Pseudouridine in HeLa rRNA

Results described in this thesis confirm and extend those of the only previous study of pseudouridine in HeLa rRNA by Amaldi and Attardi (1968). These authors estimated the pseudouridine content of rRNA by column chromatography of alkaline hydrolysates. Estimates of the number of pseudouridines were similar to, but slightly lower than, those presented in Chapter 3. These slight differences may be accounted for by possible small losses of material during repeated column chromatography. The method of separation of  $\psi_p$  from Up described in this thesis has the advantage that material is not removed from the chromatography paper at any time.

Estimates of the pseudouridine content of HeLa 45S r pre RNA described in Chapter 3 are considerably lower than those of an earlier study (Jeanteur et al, 1968). Possible alternative

explanations of this were discussed earlier. The estimate of the  $\Psi$ p content of 32S rpre RNA presented in this thesis differs somewhat from that of Jeanteur et al (1968). However, in both studies it was found that rather less pseudouridines were present in 28S RNA than in its 32S precursor. This suggests that there may be a few pseudouridines in the transcribed spacer region of 32S RNA.

Estimation of small differences in pseudouridine content between two such large molecules as 32S and 28S RNA's is difficult. Isolation and characterization of free transcribed spacer material may be necessary before accurate quantitation of pseudouridine is possible. Spot by spot comparison of  $^{14}\text{CH}_3$  fingerprints of rRNA and rpre RNA showed that 2'-O-methyl groups are absent from the transcribed spacer regions of both 32S and 45S RNA's (Maden & Salim, 1974).

Amaldi and Attardi (1968) concluded that pseudouridine occurs at specific sites in rRNA and is not spread throughout all "uridine" containing sequences. This conclusion was based on the fact that the ratio of  $\Psi$ p containing sequences to homologous Up containing sequences varied considerably between several short oligonucleotides. These authors did not investigate the presence or absence of pseudouridine in longer, unique oligonucleotides. Results described in Chapter 4 confirm that pseudouridine is present in small, but

differing amounts in various short oligonucleotides of high molar yield. In addition, the pseudouridine content of longer, unique oligonucleotides was investigated. It was found that  $\Psi$ p is present as a high percentage of Up plus  $\Psi$ p in certain of these longer sequences but was absent from others. The amount of  $\Psi$ p found in some sequences suggests that modification of Up to  $\Psi$ p is frequently an "all or nonevent". Amaldi and Attardi were unable to determine this as they investigated only short sequences, of high molar frequency, where the percentage of  $\Psi$ p in Up plus  $\Psi$ p is low.

## 6.2 Pseudouridine and 2'-O-methylation in rRNA

As discussed in Chapter 3 the number of pseudouridine residues in HeLa rRNA is very similar, if not identical, to the number of 2'-O-methyl groups. Three possible explanations of this phenomenon are:-

- (1) This is a chance numerical correlation and no relationship exists between the two modifications.
- (2) The two modifications are invariably associated with each other.
- (3) Pseudouridine and 2'-O-methyl groups are frequently, although not invariably, linked in rRNA.

Data on methylation and pseudouridylation in other eukaryotic species are incomplete. It does, however, appear that the two modifications are present in at least approximately equal numbers in rRNA of HeLa and L-cells and

in 28S RNA of *Xenopus*. 18S RNA of *Xenopus* apparently contains an appreciable excess of pseudouridine over 2'-0-methylation. *Dictyostelium* rRNA is markedly unlike the rRNA of the three vertebrate species in both base composition and in pseudouridine content. The exact number of 2'-0-methyl groups is unknown. It does, however, appear that both pseudouridine and 2'-0-methylation are markedly lower in *Dictyostelium* than in vertebrate species. It seems unlikely that a chance numerical correlation between the two modifications would be true for species as widely separated in evolution as man and slime mould. Pseudouridine and 2'-0-methyl groups are present in similar numbers in rRNA of *E. coli* (noted earlier). The numerical correlation between the two modifications may, therefore, be true of prokaryotes as well as eukaryotes.

If the numerical correlation between pseudouridine and 2'-0-methylation is not a chance one, it seems plausible to infer that related structural features within the rRNA sequence may give rise, frequently if not invariably, to both types of modification. Sequence studies on the rRNA of several eukaryotic species have brought to light a number of instances where the two types of modification are found in the same oligonucleotide (Maden & Salim, 1974; Khan & Maden, 1976a; Klootwijk & Planta, 1974; Nazar et al 1974, 1975).

There are, in fact, instances where both 2'-O-methylation and pseudouridylation occur in a single mononucleotide (Maden et al, 1974 -  $\psi_m$  - G; Gray, 1974 -  $\psi_m$  - A). On the other hand, several quite long oligonucleotides have been fully or partly sequenced and found to contain one type of modification, but not the other. However, the sequence of a number of nucleotides on either side of a site of modification will be necessary before any firm conclusion can be drawn. Even if 2'-O-methyl groups and  $\psi_p$  residues are linked in some way they may not be immediately adjacent in the rRNA sequence, as illustrated by data for 5.8S RNA.

Extensive data on the sequence of nucleotides adjacent to secondary modifications are available for 5.8S rRNA. Fig. 6.1 (a) shows a part of the rat hepatoma 5.8S rRNA sequence written as an "arm" with a stem and loop (Nazar et al., 1975b; 1976). There is a  $\psi_p$  near one end of the stem and a 2'-O-methyl group in the loop. The yeast 5.8S sequence contains a  $\psi_p$  at a corresponding site in the proposed secondary structure, but the loop is shorter and unmethylated (Fig. 6.1 (b)) (Rubin, 1973; secondary structure according to Nazar et al, 1975b, 1976). This suggests that in 5.8S rRNA at least, pseudouridine and methylation are sometimes, although not invariably, associated. 5.8S RNA, like the 18S and major 28S sequences, is transcribed and 2'-O-methylated within



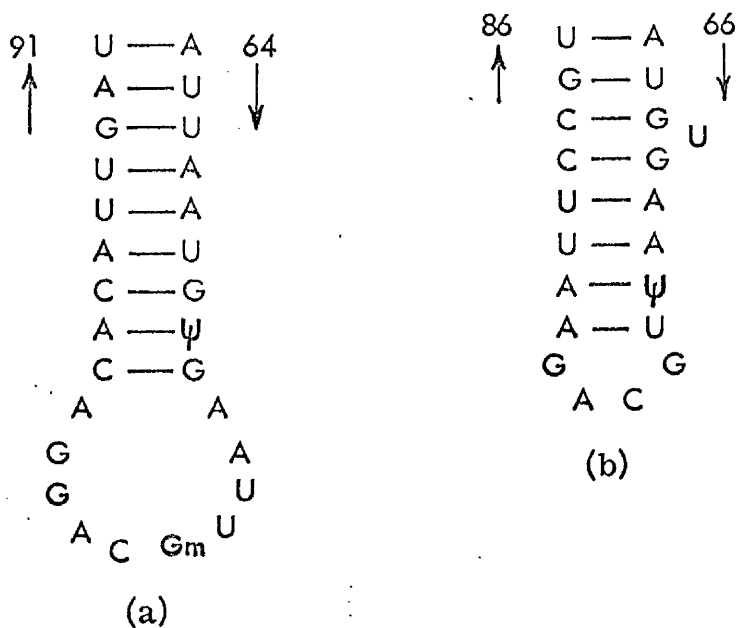


Fig. 6.1 An "arm" in the proposed secondary structure of rat hepatoma (a) and yeast (b) 5.8S RNA's. Numbers denote nucleotides from the 5' end.

45S RNA (Maden & Robertson, 1974; Nazar et al., 1975b). It will be interesting to discover, as more sequence data become available, whether arrangements such as that in Fig. 6.1a also occur in high molecular weight rRNA's.

### 6.3 Modified Nucleotides and RNA structure

Studies by Pochon et al (1964) have shown that poly pseudouridylic acid forms stable secondary interactions both with itself and with other polynucleotides. Polyuridylic acid, in contrast, does not form stable secondary interactions. The reason for the extra ability of poly  $\psi$  to form stable interactions is unclear. However, the presence of a second -NH group in pseudouridine provides additional hydrogen bonding capacity and this may well contribute to the stability of secondary interactions.

rRNA McLennan and Lane (1968) studied the digestion of wheat rRNA with a preparation of snake venom phosphodiesterase which also contained endonucleolytic activity. An unexpectedly high proportion of  $\psi$ p was found among the 3'-termini of the fragments produced. It was concluded that endonucleolytic scissions occur preferentially at pseudouridines and (or) that exonucleolytic attack is arrested by  $\psi$ p residues. Studies described here (Chapter 5) and elsewhere (e.g. Fellner, 1974) have shown that regions of strong secondary interactions in RNA are resistant to nuclease digestion. This fact and the ability of

$\Psi$ p to form strong secondary interactions lend support to the idea that  $\Psi$ p tends to arrest exonucleolytic attack. Both of the suggestions put forward by McLennan and Lane to explain their results are consistent with the notion that pseudouridine may occur at junctions between double - and single-stranded regions. in polynucleotides. These same authors also studied the proportion of methylated dinucleotides which remain intact at various stages of phosphodiesterase digestion of rRNA. They found that this proportion was roughly in parallel with the proportion of undigested RNA. 2'-O-methylation had previously been shown to inhibit hydrolysis (Gray & Lane, 1967). McLennan & Lane suggested that if 2'-O-methyl groups were situated in regions of low secondary structure, this would enhance hydrolysis sufficiently to make up for the inhibitory effect of 2-O-methyl groups. Brahms & Saudron (1966) pointed out that 2'-O-methylation would abolish hydrogen bonding between 2'-OH of ribose and oxygen of internucleotide phosphate. This could lead to reduction in the amount of secondary structure around 2'-O-methylated nucleotides and increase their susceptibility to hydrolysis.

The suggestions of McLennan & Lane upon the position of  $\Psi$ p and 2'-O-methylation in rRNA were based on indirect evidence. They do, however, seem reasonable in the light of more recent sequence data already described.

tRNA Both pseudouridine and 2'-O-methyl groups occur in tRNA molecules. The complete sequence of a considerable number of tRNA molecules is known (Barrell & Clark, 1974). Pseudouridine is found in an identical position in the TΨC' loop of almost all tRNA's. In addition, pseudouridine is frequently found elsewhere in the tRNA structure. Additional pseudouridines are most commonly found in the anticodon stem and loop, particularly in positions close to stem/loop junctions. 2'-O-methylations are much less common in published tRNA sequences than base methylations. 2'-O-methyl groups are found most commonly in the D-loop and the anticodon loop (Barrell & Clark, 1974). Dispositions of modified nucleotides similar to that found in mammalian 5.8S rRNA (Fig 6.1a) are found in several tRNA's. Pseudouridine and 2'-O-methyl groups are not, however, always found in close proximity in tRNA sequences.

A model of the tertiary structure of yeast tRNA<sup>Phe</sup> has been proposed on the basis of X-ray crystallographic data (Robertus et al, 1974; Ladner et al, 1975). The most prominent feature of this model is the presence of three major double-helical regions. These are (1) the amino acid and TΨC' stems stacking on top of each other; (2) the augmented D helix, which consists of the D stem, part of the D loop and part of the extra loop; (3) the anticodon stem, tilted at an angle to the D stem and possibly hinged to it. To these major features are joined other parts of

the molecule which have functional roles.

The T $\Psi$ C loop is involved in base-stacking interactions both within itself and with residues in the D-loop. The sequence G -T- $\Psi$ -C-G, which has been implicated in binding to the ribosome through interaction with 5S RNA (discussed in Chapter 1), is tightly tethered in the tertiary structure. Some opening out of the tertiary structure would be necessary before such an interaction could take place. It appears that the base of the pseudouridine in this sequence is involved in only two hydrogen bonds in the tertiary structure. Robertus et al (1974) have suggested that the two remaining potential H-bonding groups may interact with the methyl group of the adjacent thymidine to help form some signal for a preliminary binding step.

Yeast t-RNA<sup>Phe</sup> contains, in addition to the pseudouridine in the T $\Psi$ C loop, a second pseudouridine, which is situated at the junction between the anticodon stem and loop. The anticodon loop also contains two 2'-O-methylated nucleotides. This loop is stacked on the 3' side of the anticodon stem in an apparently exposed position. It is suggested that the join between the anticodon stem and augmented D-helix may function as a hinge and that the conformation of the anticodon loop may change during protein synthesis (Robertus et al 1974).

Determination of the tertiary structure of high molecular weight rRNA's by X-ray crystallography is at present

impossible. It is, however, possible that the tertiary structure of 5.8S RNA, which is about twice the size of tRNA, may be determined in this way.

#### 6.4. Other Possible Methods of Detection of Pseudouridine in rRNA

Methyl groups in rRNA have been studied by specific labelling with  $^{14}\text{C H}_3$  - methionine followed by fingerprinting analysis. The study of pseudouridine in rRNA has been hampered, as previously discussed, by the lack of a similar specific label. Cortese et al (1974) made use of the fact that when Up is converted to  $\Psi$ p a proton is released from the 5-position of uracil (see fig. 1.3). These authors grew bacteria in the presence of 5- $^3\text{H}$ -uridine and were then able to monitor the conversion of Up to  $\Psi$ p in tRNA in vitro by measurement of proton release from uracil. This proton release assay may have considerable application in future studies on the enzymic conversion of uridine to pseudouridine. Structural studies such as those described in this thesis would, however, seem to require the positive attachment of a label to pseudouridine rather than loss of some radioactive species.

In the absence of a direct radioactive label, location of pseudouridine may be attempted by use of the unusual chemical properties of the nucleoside. Three reactions which have been exploited in the study of pseudouridine in tRNA are

photolysis, periodate oxidation and cyanoethylation with acrylonitrile.

Photolysis. Pseudouridylate residues undergo photochemical degradation when irradiated with ultraviolet light at 257 nm (reviewed by Chambers, 1966). At the wavelength and doses employed ( $10^5$ - $10^6$  ergs/mm<sup>2</sup>), purines are largely unchanged whilst cytosine and uridine undergo reversible hydration. Tomasz & Chambers (1966) studied the photolytic degradation of pseudouridine in tRNA. They found that breakage of the polynucleotide chain at the two pseudouridines present was incomplete. A mixture of fragments was obtained which contained the products of a single cleavage at one or other pseudouridine, as well as the products of cleavage at both pseudouridines. Separation of this relatively simple mixture was carried out. However, photolysis of rRNA would yield a much larger number of products as many more pseudouridines are available for cleavage. Separation and analysis of such a complex mixture of products would be extremely difficult, particularly if the reaction does not go to completion. Because of this, photolysis of pseudouridine may not be of practical use in the study of  $\psi$ p in rRNA.

#### Periodate oxidation

The presence of a C-C glycosyl band in pseudouridine confers upon it an unusual reactivity with periodate. Oxidation of

pseudouridine by periodate follows a different reaction pathway to that of other nucleosides and is much more rapid (mechanism discussed by Chambers, 1966). However, pseudouridine occurs as only a small percentage of the total nucleosides in rRNA. Because of this, oxidation of other nucleosides constitutes a serious side reaction and probably precludes the use of selective cleavage at  $\psi$ p residues in analysis of rRNA.

### Cyanoethylation

Reaction of pseudouridine with acrylonitrile results in the addition of a cyanoethyl group to  $N_1$  of uracil. The only other nucleotides which react readily with this reagent are 4-thiouridine and inosine, neither of which is found in rRNA (Review by Ofengand, 1971). Reaction of acrylonitrile with uridine occurs at only 3% of the rate with pseudouridine. Reaction with other major nucleotides is considerably slower than with uridine. The rate of cyanoethylation in RNA varies markedly with the degree of secondary structure present. Some nucleotides which are unreactive in "native" tRNA become reactive when tRNA is denatured. The reason for this may be: (a) some residues are "buried" in the tRNA secondary structure; (b)  $N_1$  of pseudouridine may be involved in hydrogen bonding in tRNA and thus be less reactive to acrylonitrile; (c) pK of ionization of pyrimidine ring protons is higher in the polynucleotide chain than in the free nucleotide and this decreases the rate of reaction with



acrylonitrile.

The chromatographic mobility of cyanoethyl pseudouridine in several solvents is greater than that of unmodified pseudouridine but less than that of modified uridine (Ofengand, 1971). Cyanoethylated uridine has a markedly greater chromatographic mobility than the unmodified nucleotide. This means that cyanoethyl pseudouridine can be readily separated from the product of any side reaction with uridine but is less easily separated from unmodified uridine.

$^{14}\text{C}$ -acrylonitrile is commercially available. However, radiation-induced free radical polymerization, which can occur rapidly, limits the available specific activity to about 0.5mCi/m mole.  $^{14}\text{C}$ -methyl methionine used in labelling of 2'-O-methyl groups in rRNA has a specific activity approximately one hundred times greater than this.

Cyanoethylation appears to be a potentially useful technique for studying pseudouridine in rRNA. However, detection of cyanoethylated oligonucleotides in nuclease digests of rRNA may not be simple. The electrophoretic mobility of cyanoethyl pseudouridine is possibly intermediate between Up and  $\bar{\text{U}}\text{p}$ , as is its chromatographic mobility. Homologous oligonucleotides containing Up and  $\bar{\text{U}}\text{p}$  are not well resolved in fingerprints. It seems therefore, unlikely that cyanoethylated oligonucleotides will be well separated from unmodified sequences. Location

of cyanoethylated oligonucleotides in fingerprints of material labelled with  $^{14}\text{C}$ -acrylonitrile would be difficult because of the low specific activity of the label. Autoradiography of such oligonucleotides would be a lengthy process. This problem might be overcome by the addition of a small amount of  $^{32}\text{P}$  labelled RNA to aid detection of oligonucleotides by autoradiography.  $^{14}\text{C}$ -containing spots could then be identified by scintillation counting.

#### 6. 5. The Molecular Basis of Secondary Modification of rRNA

There are over two hundred modified nucleotides in HeLa cell 45S RNA. The molecular basis of this abundant secondary modification is unknown. It seems unlikely that there are over two hundred modifying enzymes, each with unique sequence specificity. An alternative possibility is that only a few enzymes are responsible and that recognition is based on features of RNA conformation. For example, one enzyme might be responsible for ribose methylation and one, or a few, for pseudouridylation. In this case, specific secondary structural features would be recognized by each modifying enzyme (see section 6.4, above, for discussion).

It is interesting that 2'-O-methyl groups are present in both tRNA and rRNA but are absent from mRNA, except for the "cap" at the 5' end of eukaryotic messengers. The presence of these secondary modifications at many sites appears to be necessary

in RNA's which form part of the translational machinery (particularly in eukaryotes). Pseudouridines and internally located 2'-O-methylations are, however, either unnecessary for messenger function or incompatible with it. Enzymes which carry out secondary modifications in the nucleus must, therefore, have some means of differentiating between RNA types.

In order to investigate the enzymology of methylation and pseudouridylation in vitro, it is necessary to obtain a template of RNA which lacks at least some of its normal secondary modifications. Vaughan et al (1967) showed that in HeLa cells deprived of methionine, 45S rpre RNA is formed which lacks many of its normal 2'-O-methyl groups. This 45S RNA is cleaved to 32S RNA but little mature cytoplasmic rRNA is formed. The large number of potential methylation sites within the 45S sequence makes difficult the use of undermethylated 45 S RNA for studies on the enzymology of methylation. There is no known method of blocking the conversion of Up to  $\Psi$ p in rRNA. Cortese et al (1974) obtained tRNA from mutants of *S. typhimurium* and *E. coli* which lacked two pseudouridines found in tRNA from WT bacteria. They were able to pseudouridylate this undermodified tRNA in vitro using an extract from WT bacteria. No mutants which produce undermodified rRNA are at present available.

Johnson and Söll (1970) synthesized unmodified tRNA in vitro from a DNA template enriched for tRNA genes. Ribosomal genes are much greater in size and their complete

in vitro transcription much more difficult. A possible alternative approach would be to use restriction endonucleases to produce defined fragments of rDNA. In vitro transcription of unmodified rRNA fragments might then be possible by the use of RNA polymerase. Such fragments would contain fewer potential modification sites than, for instance, 45S RNA from methionine-starved cells. This should make them more suitable for studies on the enzymology of secondary modification.

#### 6. 6. The Function of Secondary Modifications in rRNA

Vaughan et al (1967) showed that methylation of rRNA is essential for ribosome maturation. It is not known whether pseudouridine formation is also essential. No function has yet been ascribed to either type of modification.

It seems that only few pseudouridines and 2'-O-methylations are present in highly G-C rich regions of HeLa 28S rRNA (as described in Chapter 5). Other studies have shown that extensive G-C rich regions are present in rRNA of other vertebrate species (e.g. Cox et al, 1973, 1974; Wellauer & Dawid, 1974; Wellauer et al, 1974). The size of these G-C rich regions decreases from higher to lower vertebrates (Wellauer et al, 1974; Schibler et al, 1975). In lower eukaryotes A-U rich regions appear to replace G-C rich regions (Cox et al, 1976b). It would be interesting to determine whether these regions of atypical nucleotide composition contain low numbers of secondary modifications in all eukaryotes. If this is so,

2'-O-methyl groups and pseudouridines must be largely confined to regions of rRNA which lack large secondary-structural loops and which have base composition of approximately 50% G+C. There is considerable conservation of regions of this type between eukaryotic species (discussed in Chapter 1). Such regions may contain short helical regions interspersed with single stranded regions (as suggested by McLennan & Lane, 1968; discussed in section 6.3). It has already been shown that many distinctive methylation sites are conserved between species as widely separated in evolution as man (HeLa cells, Maden & Salim, 1974), fruit flies (Maden & Tartof, 1974) and yeast (Klootwijk & Planta, 1974). If, as seems possible, pseudouridines and 2'-O-methylations are in some way associated, many sites of pseudouridylation may also be conserved.

The conservation of 2'-O-methyl groups (and possibly pseudouridines) between widely separated eukaryotic species suggests that secondary modifications are necessary for essential functions of the eukaryotic ribosome. 2'-O-methyl groups and pseudouridines are much less common in rRNA of prokaryotes than in eukaryotic rRNA. Any function which they perform must be largely confined to eukaryotic ribosomes. The ribosomes of eukaryotes and prokaryotes perform essentially the same function in protein synthesis. Further structural and metabolic studies will be necessary before the differences between 80S and 70S ribosomes can be explained.

The study of pseudouridine in eukaryotic rRNA is still at an early stage of development. More detailed analysis of the nucleotide sequences around sites of pseudouridylation, the relationship between pseudouridylation and methylation, and the molecular basis of pseudouridine formation may have an important bearing on understanding the distinctive structural properties of eukaryotic rRNA's.

Appendix

During the early part of my research studentship I carried out studies on the frequency of short oligonucleotide sequences within Poliomyelitis and Encephalomyocarditis Virus RNA's. In addition to gaining data on an interesting subject, this investigation served to familiarise me with many techniques which were later used in my major research project. The background to this research and the results obtained are described in the attached publication.

Amaldi, F. & Attardi, G. (1968). J. Mol. Biol. 32, 737-755.

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

Barrell, B.G. & Clark, B.F.C. (1974). Handbook of Nucleic  
Acid Sequences. Joynson & Bruvvers, Oxford.

Birnstiel, M.L., Chipchase, M. & Spiers, J. (1971).

in "Progress in Nucleic Acid Research and Molecular  
Biology" 11, 33-79.

(Davidson, J.N & Cohn, W.E. Eds.). Academic Press, New York.

Birnstiel, M.L. & Grunstein (1972). FEBS Symp. 23, 349-365.

Birnstiel, M.L., Spiers, J., Purdon, I., Jones, K. and Loening U.E.

(1968), Nature 219, 454-463.

\*

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

Brown, D.D & Gurdon, J.M. (1964) Proc. Natl. Acad. Sci. U.S.A.  
51, 139-146.

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

Brownlee, G.G. (1972). Lab. Techniques in Biochem. and Mol. Biol.

Determination of Sequences in RNA (Work, T.S & Work, E. Eds.)

North Holland Publishing Co., Netherlands.

Brownlee, G.G. & Sanger, F. (1967). J. Mol. Biol. 23, 337-353.

Burdon, R.H. (1975). in "Processing of RNA", pp 138-153,

Brookhaven Symposium no. 26.

Chambers, R.W. (1966). Prog. Nucl. Acid Res. 5, 349-398.

Brahms, J. & Saudron, C. (1966) Nature 212, 1309-1312.

\*



- Claude, A. (1941). Cold Spr. Harb. Symp. Quant. Biol.  
9, 263-271.
- Cohn, W.E. (1957). Federation Proc. 16, 166.
- Cohn, W.E. (1960). J. Biol. Chem. 235, 1488-1498.
- Cortese, R., Kammen, H.O., Spengler, S.J. and Ames, B.N. (1974).  
J. Biol. Chem. 249, 1103-1108.
- Cox, R.A., (1966). Biochem J. 98, 841-857.
- Cox, R.A. (1970). Biochem. J. 117, 101-118.
- Cox, R.A., Hirst, W., Godwin, E. and Kaiser, I. (1976a).  
Biochem. J. 155, 279-295.
- Cox, R.A., Godwin, E. & Hastings, J.B. (1976b). Biochem.J.  
155, 465-475.
- Cox, R.A., Huvös, R., and Godwin, E. (1973). Israel J. of Chemistry  
11, 407-422.
- Cox, R.A., Pratt, H., Huvös, P., Higginson, B., and Hirst W.  
(1973b). Biochem. J. 134, 775-793.
- Dahlberg, J.E., Nickolaev, N. and Schlesinger, D. (1975) in  
Processing of RNA", pp. 194-200, Brookhaven Symposium No. 26.
- Dasgupta, R., Shih, D.S., Caris, C. and Kaesberg, P. (1975).  
Nature 256, 624-628.
- Davis, F.F. and Allen, F.W. (1957). J. Biol. Chem. 227, 907-915.
- Davis, R.W., Simon, M. and Davidson, N. (1971). Methods in  
Enzymology, 21, 413-428.
- Dawid, I.B., Brown, D.D. and Reeder, R.H. (1970). J. Mol. Biol.  
51, 341-360.

- Dawid, I.B. and Chase, J.W. (1972). J. Mol. Biol. 63, 217-231.
- Dawid, I.B. and Wellauer, P.K. (1976). Cell 8, 443-448.
- Delauney, J., Creusot, F. and Schapira, G. (1973).  
Eur. J. Biochem. 39, 305-312.
- De Wachter, R. and Fiers, W. (1971). Methods in Enzymology 21,  
167-178.
- Dubin, D.T. and Gunlap, A. (1967). Biochim. Biophys. Acta., 134,  
106-123.
- Dunn, D.B. (1959). Biochim. Biophys. Acta. 34, 286.
- Eagle, H. (1959). Science 130, 432-437.
- Earle, W.R. (1943). J. Natl. Cancer Inst. 4, 165-212.
- Ehresmann, C., Stiegler, P., Fellner, P. and Ebel, J.P. (1972).  
Biochimie 54, 901-967.
- Fellner, P. (1969). Eur. J. Biochem. 11, 12-27.
- Fellner, P. (1974). in "Ribosomes" pp. 169-191 (Nomura, M.,  
Tissieres, A. and Lengyel, P. eds.). Cold Spring Harbor  
Press, Long Island, New York.
- Firtel, R.A. and Lodish H. (1973). J. Mol. Biol. 79, 295-314.
- Fraser, N.W., Maden, B.E.H. and Burdon, R.H., (1973).  
FEBS Letts. 36, 257-260.
- Fuller, W., Hutchinson, F., Spencer, M. and Wilkins, M.H.F. (1967)  
J. Mol. Biol. 27, 507-524.
- Gall, J.G. (1968). Proc. Natl. Acad. Sci. U.S.A. 60, 553-560.

Gey, G. O., Coffman, W. D. and Kibicek, M. T. (1952).

Cancer Res. 12, 264-265.

Girard, M., Penman, S. and Darnell, J. E. (1964).

Proc. Natl. Acad. Sci. U.S.A. 51, 205-211.

Glitz, D. G. and Decker, C. A. (1963). Biochemistry, 2, 1185-1192.

Goddard, J. P. and Maden, B. E. H. (1976). Nucl. Acids Res.

3, 431-440.

Goddard, J. P. and Schulman, L. H. (1972). J. Biol. Chem.

247, 3864-3867.

Godwin, E., Cox, R. A. and Hüvös, P. (1974). Acta. Biol.

Med. Germ. 33, 733-752.

Goldwasser, E. and Heinrickson, R. L. (1966). Progress Nucl.

Acid Res. 5, 399-416.

Gould, H. J. (1966). Biochemistry, 5, 1103-1108.

Gould, H. J. (1967). J. Mol. Biol. 29, 307-313.

Gould, H. J., Bonanou, S. and Kanalingham, K. (1966).

J. Mol. Biol. 22, 397-399.

Gray, M. W. (1974). Biochemistry 13, 5453-5463.

Gray, M. W. and Lane, B. G. (1967). Biochim. Biophys. Acta.

134, 243-257.

Greenberg, J. R. and Penman, S. (1966). J. Mol. Biol.

23, 149-161.

Hackett, P. B. and Sauerbier, W. (1975). J. Mol. Biol.

91, 235-256.

- Hadjilov, A. A. and Cox, R. A. (1973). Biochem. J. 135 349-351.
- Hashimoto, S., Sakai, M. and Muramatsu, M. (1975). Biochemistry 14, 1956-1964.
- House, W. and Wildy, P. (1965). Lab. Practice 14, 594-595.
- Jacobson, A., Firtel, R. H. and Lodish, H. F. (1974). J. Mol. Biol. 82, 213-230.
- Jeanteur, P., Amaldi, F. and Attardi, G. (1968). J. Mol. Biol. 33, 755-775.
- Jeanteur, P. and Attardi, G. (1969) J. Mol. Biol. 45, 305-324.
- Jeneer, R. and Brachet, J. (1941). Acta. Biol. Belg. 1, 476.
- Jeneer, R. and Brachet, J. (1942). Acta. Biol. Belg. 2, 447.
- Johnson, L. and Soll, D. (1970). Proc. Natl. Acad. Sci. U.S.A. 67, 943-950.
- Jordan, B. R. (1971). J. Mol. Biol. 55, 423-439.
- Kabat, D. (1975). J. Biol. Chem. 250, 6085-6092.
- Kanamaru, R., Choi, Y. C. and Busch, H. (1974). J. Biol. Chem. 249, 2453-2465.
- Khan, M. S. N. and Maden, B. E. H. (1976a). J. Mol. Biol. 101, 235-254.
- Khan, M. S. N. and Maden, B. E. H. (1976b). FEBS. Letters 61, 10-13.
- Klootwijk, J. and Planta, R. J. (1973). Eur. J. Biochem. 39, 325-333.
- Klootwijk, J. and Planta, R. J. (1974). Mol. Biol. Rep. 1, 187-191.

Krystosek, A., Bitte, L.F., Cawthorn, M.L. and Kabat, D.

(1974) in "Ribosomes" p.p. 855-870 (Nomura, M.,  
Tissieres, A. and Lengyel, P, eds.), Cold Spring Harbor  
Press, Long Island, New York.

Kumar, A. and Warner, J.R. (1972). J. Mol. Biol. 63, 233-246.

Ladner, J.E., Jack, A., Robertus, J.D., Brown, R.S., Rhodes, D.,  
Clark, B.F.C. and Klug, A. (1975). Proc. Natl. Acad. Sci.  
U.S.A. 72, 4414-4418.

Lane, B.G. (1965). Biochemistry, 4, 212-219.

Lane, B.G. and Tamaoki, T. (1969). Biochim. Biophys. Acta.  
179, 332-340.

Lau, R.Y., Kennedy, T.D. and Lane, B.G. (1974). Can. J.  
Biochem. 52, 1110-1123.

Liau, M.C. and Hurlbet, R.B. (1975). J. Mol. Biol.  
98, 321-332.

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

Mackie, G.A. and Zimmerman, R.A. (1975). J. Biol. Chem.  
250, 4100-4112.

Maden, B.E.H. (1971) in "Progress in Biophysics and Molecular  
Biology" 22, 129-177 (Butler, J.A.V & Noble, D. eds.),  
Pergamon Press, New York.

Maden, B.E.H. (1976). TIBS 1, 196-199.

Maden, B.E.H. and Forbes, J.S. (1972). FEBS Letts. 28, 289-292.

- Maden, B. E. H., Forbes, J. S., de Jonge, P., and Klootwijk, J.  
(1975). FEBS Letts. 59, 60-63.
- Maden, B. E. H. and Robertson, J. S. (1974). J. Mol. Biol.  
87, 227-235.
- Maden, B. E. H., Salim, M. and Summers, D. F. (1972). Nature  
New Biol. 237, 5-9.
- Maden, B. E. H. and Salim, M. (1974). J. Mol. Biol. 88, 133-164.
- Maden, B. E. H. and Tartof, K. (1974). J. Mol. Biol. 90, 51-64.
- Markham, R. and Smith, J. D. (1952). Biochem. J. 52, 552-557.
- McLennan, B. D. and Lane, B. G. (1968). Can. J. Biochem. 46, 81-92.
- McPhie, P. and Gratzner, W. B. (1966). Biochemistry 5, 1310-1315.
- Michelson, A. M. and Cohn, W. E. (1962). Biochemistry 1, 490-495.
- Monier, R. (1974) in "Ribosomes", p.p. 141-168 (Nomura, M.,  
Tissières, A. and Lengyel, P., eds.). Cold Spring Harbor  
Press, Long Island, New York.
- Mowshowitz, D. B. (1970). J. Mol. Biol. 50, 143-151.
- Nazar, R. N., Sitz, T. O. and Busch, H. (1975a). J. Biochem.  
250, 8591-8597.
- Nazar, R. N., Sitz, T. O. and Busch, H. (1975b). Biochem. Biophys.  
Res. Commun. 62, 736-743.
- Nazar, R. N., Sitz, T. O. and Busch, H. (1976). Biochemistry,  
15, 505-508.
- Nichols, J. L. and Lane, B. G. (1967). J. Mol. Biol. 30, 477-489.

Nishikawa, K. and Takemura, S. (1974). FEBS Letts. 40, 106-109.

Nishimura, S. (1972). Progr. Nucl. Acid. Res. Mol. Biol. 12, 50-86.

Noller, H. F. (1974). Biochemistry 13, 4694-4703.

\*

Pene, J. J., Knight, E. and Darnell, J. E. (1968). J. Mol. Biol.

33, 609-623.

Penman, S. (1966). J. Mol. Biol. 17, 117-130.

Penman, S. (1969). in "Fundamental Techniques in Virology"

(Habel, K. & Salzman, N. P. eds.). Chapter 5, Academic

Press, New York.

Penman, S., Smith, I. and Holtzman, E. (1966). Science, 154, 786-789

Perry, R. P. and Kelly, D. E. (1972). J. Mol. Biol. 70, 265-279.

Petermann, M. L. and Pavlovec, A. (1966). Biochim. Biophys. Acta. 299,

173-175.

Pinder, J. C., Gould, H. J. and Smith, I. (1969). J. Mol. Biol.

40, 289-298.

Pochon, F. D., Michelson, A. M., Grunberg-Manago, M., Cohn, W. E.

and Dondon, L. (1964). Biochim. Biophys. Acta. 80, 441-447.

Reeder, R. H. (1974). in "Ribosomes" p. p. 489-518 (Nomura, M.,

Tissieres, A. and Lengyel, P. Eds.). Cold Spring Harbor

Press, Long Island, New York.

Reeder, R. H., Higashimakagawa, T. and Miller, O. (1976).

Cell 8, 449-454.

Ritossa, F. M. and Spiegelman, S. (1965). Proc. Natl. Acad. Sci.

U. S. A. 53, 737-745.

Robberson, D. L., Aloni, Y., Attardi, G., and Davidson, N.

(1971). J. Mol. Biol. 60, 473-484.

\*

Robertson, J. S. and Maden, B. E. H. (1973). *Biochim. Biophys. Acta.*

331, 61-70.

Robertus, J. D., Ladner, J. F., Finch, J. T., Rhodes, D., Brown, R. S.,

Clark, B. F. C. and Klug, A. (1974). *Nature* 250, 546-551.

Rubin, G. M. (1973). *J. Biol. Chem.* 248, 3860-3875.

Sanford, K. K., Earle, W. R. and Likley, G. D. (1948).

*J. Natl. Cancer Inst.* 9, 229-246.

Sanger, F., Brownlee, G. G. and Barrell, B. G. (1965).

*J. Mol. Biol.* 13, 373-398.

Saponara, A. G. and Enger, M. D. (1974). *Biochim. Biophys. Acta.*

349, 61-77.

Scherrer, K. and Darnell, J. E. (1962). *Biochim. Biophys. Res.*

*Commun.* 7, 486-490.

Scherrer, K., Latham, H. and Darnell, J. E. (1963). *Proc. Natl.*

*Acad. Sci. U.S.A.* 49, 240-248.

Schibler, U., Wyler, T. and Hagenbüchle, O. (1975).

94, 503-517.

Shepherd, J. and Maden, B. E. H. (1972). *Nature* 236, 211-214.

Simsek, M., Petrissant, G. and Raj Bahandary, U. L. (1973).

*Proc. Natl. Acad. Sci. U.S.A.* 70, 2600-2604.

Sinclair, J. and Brown, D. D. (1971). *Biochemistry*, 10, 2761-2769.

Singh, H. and Lane, B. G. (1964). *Canad. J. Biochem.* 42, 87-93.

Soeiro, R., Vaughan, M. H. and Darnell, J. E. (1968). *J. Cell Biol.*

36, 91-101.



- Spiers, J. and Birnstiel, M. L. (1974). J. Mol. Biol. 87, 23.
- Steitz, J. A. and Jakes, K. (1975). Proc. Natl. Acad. Sci. U.S.A. 72, 4734-4738.
- Tinoco, I., Uhlenbeck, O. C. and Levine, M. D. (1971). Nature 230, 362-367.
- Tischendorf, G. W., Zeichardt, H. and Stöffler, G. (1975). Proc. Natl. Acad. Sci. U.S.A. 72, 4820-4824.
- Tissieres, A. (1974). in "Ribosomes", p.p. 3-12 (Nomura, M., Tissieres, A. and Lengyel, P. eds.). Cold Spring Harbor Press, Long Island, New York.
- Tomasz, M. and Chambers, R. W. (1966). Biochemistry 5, 773-782.
- Trapman, J., De Jonge, P. and Planta, R. J. (1975). FEBS Letts. 57, 26-30.
- Udem, S. A. and Warner, J. R. (1972). J. Mol. Biol. 65, 227-242.
- Van den Bos, R. C. Retel, J. and Planta, R. J. (1971). Biochim Biophys. Acta. 232, 494-508.
- Vaughan, M. H., Soiero, R., Warner, J. R. and Darnell, J. E. (1967). Proc. Natl. Acad. Sci. U.S.A. 58, 1527-1534.
- Wallace, H. and Birnstiel, M. L. (1966). Biochim. Biophys. Acta. 114, 296-310.
- Warner, J. R. and Soiero, R. (1967) Proc. Natl. Acad. Sci. U.S.A. 58, 1984-1990.
- Watts, D. J. and Ashworth, J. M. (1970). Biochem. J. 119, 171-174.

Weinberg, R.A., Loening, U.E., Willems, M. and Penman, S. (1967).

Proc. Natl. Acad. Sci. U.S.A. 58, 1088-1095.

Weinberg, R.A. and Penman, S. (1968). J. Mol. Biol. 38, 289-304.

Weinberg R.A. and Penman, S. (1970). J. Mol. Biol. 47, 169-178.

Weiss, S.B. and Legault-Demare J. (1965). Science 149, 429-430.

Wellauer, P.K. and Dawid, I.B. (1973). Proc. Natl. Acad. Sci. U.S.A.

70, 2827-2831.

Wellauer, P.K. and Dawid, I.B. (1974). J. Mol. Biol. 89, 379-395.

Wellauer, P.K., Reeder, R.H., Carroll, D., Brown, D.D., Deutch, A.,

Higashinakagawa, T. and Dawid, I.B. (1974a). Proc. Natl.

Acad. Sci. U.S.A. 71, 2823-2827.

Wellauer, P.K., Dawid, I.B., Kelly, D.E. and Perry, R.P. (1974b).

J. Mol. Biol. 89, 397-407.

Wellauer, P.K. and Reeder, R.H. (1975). J. Mol. Biol. 94, 151-161.

Wikman, J., Howard, E. and Busch, H. (1969). J. Biol. Chem.

244, 5471-5480.

Williems, M., Wagner, E., Laing, R. and Penman, S. (1968)

J. Mol. Biol. 32, 211-221.

Winicov, I. (1976). J. Mol. Biol. 100, 141-155.

Winocour, E., Kay, A.M. and Stoller, V. (1965). Virology,

27, 156-169.

Woese, C.R., Fox, G.F., Zablen, L., Uchida, T., Boren, L.,

Pechman, K., Lewis, R.L. and Stahl, D. (1975).

Nature, 254, 83-86.

(Nomura, M., Tissieres, A. and Lengyel, P. eds.).

Cold Spring Harbor Press, Long Island, New York.

Wyatt, P. (1951). Biochem J. 48, 584-590.

Zimmerman, R.A. (1974) in "Ribosomes" p.p. 225-269.

(Nomura, M., Tissieres, A. and Lengyel, P. eds.).

Cold Spring Harbor Press, Long Island, New York.

Zimmerman, E. F. and Holler, B. W. (1967). J. Mol. Biol.

23, 149-161.

---

# Reprint from Nucleic Acids Research

---

Published by Information Retrieval Limited  
1 Falconberg Court London W1V 5FG England



---

Information Retrieval Ltd  
1 Falconberg Court  
London W1V 5FG England

---

## Executive editors

A.S.Jones, Birmingham  
Dieter Söll, Yale  
R.T.Walker, Birmingham  
E.S.Krudy, IRL, London

Production Editor:  
Alexandra Williamson

## Editorial board

A.A.Bayev, Moscow  
G.Bernardi, Paris  
R.H.Burdon, Glasgow  
C.Cantor, Columbia, New York  
B.F.C.Clark, Aarhus  
R.A.Cox, London  
D.M.Crothers, Yale  
F.Eckstein, Göttingen  
A.Falaschi, Pavia  
P.Fellner, High Wycombe  
M.Geffer, MIT  
P.T.Gilham, Indiana  
R.F.Goldberger, NIH  
A.F.Graham, Montreal  
R.H.Hall, Hamilton  
P.H.Hofschneider, Munich  
A.Holý, Prague  
M.Ikehara, Osaka  
P.Lengyel, Yale  
R.L.Letsinger, Evanston, Ill.  
S.Linn, Berkeley  
W.E.G.Müller, Mainz  
S.Nishimura, Tokyo  
S.Olsnes, Oslo  
U.L.RajBhandary, MIT  
J.D.Smith, Cambridge  
J.A.Steitz, Yale  
I.O.Walker, Oxford  
J.P.Waller, Gif-sur-Yvette  
S.M.Weissman, Yale  
M.Yarus, Boulder, Colorado



---

Non-Random Frequencies of Short Oligonucleotide Sequences within Poliomyelitis and Encephalomyocarditis Virus RNAs.

---

D. G. Hughes and B. E. H. Maden

---

Department of Biochemistry, University of Glasgow, Glasgow, G12 8QQ,  
U.K.

---

Received 3 June 1975

---

ABSTRACT

In poliomyelitis and encephalomyocarditis virus RNAs the relative frequencies of several short oligonucleotide sequences differ substantially from those expected on a random basis. The frequencies of CpGp-containing sequences are low, and the frequency of GpUpApGp is considerably lower than that of GpApUpGp. The relationship of these findings to those obtained from host cell RNAs is discussed.

INTRODUCTION

Nearest neighbour analysis of DNA from many sources has revealed that the vertebrate genome is specifically deficient in the doublet, CpGp<sup>1,2</sup>. The DNAs of several small mammalian viruses are also CpGp deficient, whereas various other, larger mammalian virus DNAs reveal different doublet patterns and are not CpGp deficient<sup>2,3,4</sup>. The nucleic acids of certain vertebrate RNA viruses, including encephalomyocarditis virus<sup>5</sup> (EMCV), fowl plague virus<sup>6</sup> and Newcastle disease virus<sup>7</sup>, also revealed CpGp deficiencies when examined by nearest neighbour analysis. The term, "general design", has been used to describe such non-random features of nearest neighbour relationships within nucleic acids<sup>2,8</sup>.

RNA fingerprinting methods can also provide information of this kind. T<sub>1</sub> ribonuclease fingerprints of

---

Hela cell heterogeneous nucleoplasmic RNA (HnRNA)<sup>9,10</sup> revealed CpGp deficiency comparable to that found by nearest neighbour analysis of mammalian DNA. 28 S ribosomal RNA, a transcript of a specialized and very restricted part of the Hela cell genome, did not reveal CpGp deficiency<sup>9</sup>.

In order to further investigate possible relationships between mammalian genetic material and that of the smaller mammalian RNA viruses, we have carried out a fingerprinting analysis of oligonucleotide frequencies within RNAs from two picornavirus subgroups, poliomyelitis and encephalomyocarditis viruses. This paper describes our findings.

#### METHODS

<sup>32</sup>P labelled poliovirus type 1 RNA was purified from virus particles which were extracted from infected Hela cells labelled in the presence of actinomycin D, and was a gift from Dr. D.F. Summers. <sup>32</sup>P labelled EMCV RNA was prepared according to Porter, Carey and Fellner<sup>11</sup>, and was a gift from Dr. P. Fellner. 5 x 10<sup>5</sup> c.p.m. of RNA were used in each fingerprinting analysis, in the presence of unlabelled carrier RNA added to a final concentration of 20 µg. RNA was lyophilized and digested with T<sub>1</sub> ribonuclease (Sankyo, 2 µg) in approximately 5 microlitres of 0.01 M Tris HCl (pH 7.4), 0.001 M EDTA, for 30 min at 37°. The products were separated by electrophoresis on cellulose acetate at pH 3.5 in the presence of 7 M urea for 3 hr at 4.7 KV, followed by DEAE paper in 7% formic acid for 16 hr at



1.2 KV<sup>12</sup>. Oligonucleotides were detected by autoradiography. The identities of all the smaller products (those indicated in figure 1) were known from their mobilities (refs 12, 13 and observations in this laboratory).

Molar yields of the oligonucleotides were determined by eluting the spots, counting in toluene PPO scintillant, and dividing the c.p.m. by two for dinucleotides, three for trinucleotides etc. The yields were expressed relative to an arbitrary value of 100 for Gp plus cyclic Gp, and are the means of three poliovirus determinations and two EMCV determinations (table 1).

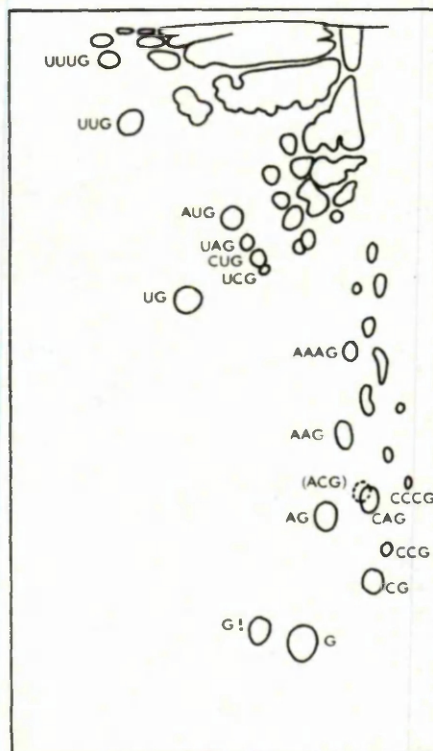
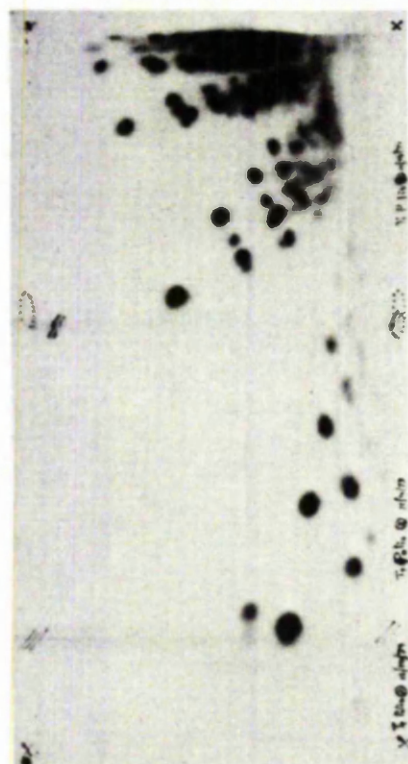
Theoretically expected molar yields (table 1, parentheses) were calculated relative to 100 for Gp on the basis of considerations outlined in the following examples.

Table 1				
Relative molar frequencies of oligonucleotides in T <sub>1</sub> ribonuclease digests of poliovirus RNA, EMCV RNA and HeLa cell HnRNA.				
RNA	Oligonucleotide			
	Gp	CpGp	CpCpGp	CpCpCpGp
Polio	100(100)	9.03(23.8)	1.18(5.66)	0.65(1.35)
EMC	100(100)	8.06(21.7)	5.52(4.71)	1.08(1.02)
HeLa Hn	100(100)	3.90(24.3)	1.26(5.9)	0.47(1.43)
	ApGp	ApApGp	ApApApGp	
Polio	22.7(29.3)	8.62(8.58)	2.53(2.52)	
EMC	24.2(27.9)	8.87(7.78)	3.20(2.17)	
HeLa Hn	31.5(27.8)	7.46(7.73)	2.65(2.15)	
	UpGp	UpUpGp	UpUpUpGp	
Polio	22.9(23.8)	7.24(5.66)	1.83(1.35)	
EMC	19.0(26.4)	3.80(6.97)	- (1.84)	
HeLa Hn	23.9(28.4)	8.06(8.06)	4.25(2.29)	
	UpCpGp	CpUpGp	UpApGp	ApUpGp
Polio	1.38(5.66)	5.36(5.66)	1.67(6.97)	10.07(6.97)
EMC	2.16(5.73)	4.28(5.73)	3.93(7.36)	8.30(7.36)
HeLa Hn	0.33(6.90)	8.07(6.90)	6.50(7.92)	8.70(7.92)

Experimentally determined yields are followed by theoretically expected yields (parentheses). Both were calculated relative to 100 for Gp as described in methods. Base compositions used for calculating theoretical values were:- for polio RNA<sup>14</sup>; Ap, 29.3%; Up, 23.8%; Gp, 23.2%; Cp, 23.8%; for EMCV RNA<sup>5</sup>; Ap, 27.9%; Up, 26.4%; Gp, 23.9%; Cp, 21.7%. An accurate experimental value was not obtained for (Up)<sub>3</sub>Gp in EMCV RNA. The HeLa HnRNA values were published previously<sup>9</sup>, except for UpApGp and ApUpGp, which were kindly provided by H.W. Fraser (personal communication).



The  $T_1$  product ApGp is derived from the sequence GpApGp. The expected frequency of ApGp, relative to XpGp, where Xp is any nucleotide, is proportional to the percentage of Ap in the molecule, or 29.3% for polio RNA<sup>14</sup>. (Base composition data are shown in the legend to table 1). The expected frequency of Gp being next to ApGp is proportional to the % Gp in the molecule. That of Gp being next to Gp (giving the  $T_1$  product Gp) is also proportional to % Gp. Therefore the expected frequency of (Gp)ApGp, relative to 100 for (Gp)Gp, is given by % A, the % G "cancelling out". Similarly the probable frequencies of (for example) ApUpGp or UpApGp relative to 100 for Gp are both given by % A x % U, or  $29.3\% \times 23.8\% = 6.97\%$  for polio RNA.



Polio RNA Key

Figure 1

Ribonuclease  $T_1$  fingerprint of  $^{32}\text{P}$  labelled poliovirus type I RNA, with Key. First dimension, right to left, cellulose acetate, pH 3.5. Second dimension, downwards, DEAE paper, 7% formic acid.

## RESULTS

Figure 1 shows a  $T_1$  ribonuclease fingerprint of poliovirus type I RNA. Molar recoveries of several short oligonucleotides were determined relative to 100 for Gp as described under "methods". Similar determinations were carried out on EMCV RNA. The results are summarized in Table 1, together with previously determined values from Hela cell HnRNA<sup>9</sup> for comparison, and, in parentheses, theoretically expected yields for the various RNAs, also calculated as described in "methods". There is good agreement between observed and expected values for several of the oligonucleotides. However, there are two major groups of discrepancies.

(i) CpGp. In polio RNA the product CpGp, derived from the sequence GpCpGp, occurs in much lower than the expected yield. This is also true for other products terminated by CpGp. In particular UpCpGp is much less abundant than CpUpGp, as is evident from visual inspection of the fingerprint (figure 1). ApCpGp and CpApGp were not quantitated because of incomplete resolution. Nevertheless visual inspection suggests "skewing" of label towards CpApGp, the isomer ApCpGp being only weakly labelled.

In EMCV RNA CpGp and UpCpGp occur in low yield, whereas (Cp)<sub>2</sub>Gp and (Cp)<sub>3</sub>Gp occur in approximately the statistically expected frequencies, though less abundant overall than CpGp. Inspection of the fingerprints suggested that ApCpGp was less abundant than CpApGp, but, as for polio RNA, these two isomers were insufficiently resolved to permit accurate quantitation.

(ii) UpApGp and ApUpGp. In poliovirus RNA the product, UpApGp, derived from the sequence, GpUpApGp, occurs in considerably lower yield than the isomeric product, ApUpGp, derived from the sequence, GpApUpGp. The difference is evident on examining the fingerprint (figure 1). A similar but less pronounced inequality occurs in EMCV RNA.

A third, possibly minor point is that in EMCV RNA the yield of UpUpGp is rather low.

### DISCUSSION

The mammalian and viral genomes differ vastly in their sequences complexities, those of mammalian DNA or Hela cell HnRNA being several orders of magnitude greater than those of polio or EMCV RNAs. Therefore, before discussing the possible relationship of these viral results to those obtained from mammalian cells, it is useful to obtain a rough estimate of the absolute abundances, within the viral RNAs, of the sequences in table 1. This is done explicitly here for polio RNA.

The molecular weight of polio RNA is approximately  $2.6 \times 10^6$  (refs 15, 16). This corresponds to some 7,800 nucleotides, of which 23.2% are Gp residues<sup>14</sup>. Therefore GpGp, giving rise to the  $T_1$  product Gp, should occur  $.232 \times .232 \times 7,800$ , or some 420 times per molecule on a random nearest neighbour basis. Thus the expected relative values in table 1 should be multiplied by 4.2 to give expected absolute frequencies. The observed values should be multiplied by a similar, but not necessarily identical factor to give observed absolute frequencies. (The real

frequency of GpGp may differ by an unknown amount from the expected frequency). In fact 4.62 was chosen as a plausible normalization factor for the observed values, as this was the nearest value to 4.2 that resulted in a whole number yield for the least abundant product, (Cp)<sub>3</sub>Gp (three moles per mole of RNA).

The recalculated values are shown in table 2.

Table 2			
Estimated observed and expected absolute frequencies of oligonucleotides in poliovirus RNA T <sub>1</sub> ribonuclease fingerprint.			
Gp	CpGp	CpCpGp	CpCpCpGp
462(420)	42(100)	5-6(24)	3(5.7)
	ApGp	ApApGp	ApApApGp
	105(123)	40(36)	11-12(10.6)
	UpGp	UpUpGp	UpUpUpGp
	106(100)	33(24)	8-9(5.7)
UpCpCp	CpUpGp	UpApGp	ApUpGp
6-7(24)	25(24)	7-8(29)	46(29)

All values are approximate, and were calculated using assumptions described in the discussion. The expected yields are in parentheses. For low yield products the observed yields are given to the nearest integers below and above the actual normalized values (e.g., an actual value of 5.5 for CpCpGp is likely to signify 5 or 6). It should be noted that all of these products are preceded by Gp in the intact RNA.

If the assumptions underlying the calculations are even approximately correct, the expected abundances of products up to and including trinucleotides (i.e. sequences up to and including tetranucleotides) are considerable. It seems reasonable to conclude that the generally low frequency of CpGp-terminated sequences, and the inequality between (Gp)ApUpGp and (Gp)UpApGp, represent real positive or negative preferences for these short sequences within the overall primary structures of the viral RNAs.

Polio and EMCV RNAs therefore appear to resemble, qualitatively, the mammalian genome and Hela cell HnRNA with respect to overall CpGp deficiency. For EMCV RNA this confirms previous findings<sup>5</sup>, but the confirmation is useful, because in the previous study by nearest neighbour analysis conditions of somewhat "forced" copying were used, employing M. lysodeikticus RNA polymerase<sup>5</sup>. The extent of CpGp deficiency observed here is not as great as in Hela cell HnRNA. Moreover HnRNA differs from these viral RNAs by showing only a minor inequality between (Gp)ApUpGp and (Gp)UpApGp (table 1).

Polio and EMCV RNAs serve as their own messengers. Only a fraction of the total sequences within HnRNA give rise to messenger sequences<sup>17</sup>. It would be of interest to know to what extent these viral RNAs resemble mammalian messenger RNAs, as opposed to HnRNA. A recent analysis of Hela cell messenger RNA by the present method (N.W. Fraser and R.H. Burdon, submitted for publication and personal communication) indicates that (i) CpGp deficiency in messenger occurs but is less extreme than in HnRNA, and (ii) messenger RNA yields a substantial excess of (Gp)ApUpGp over (Gp)UpApGp.

UpApGp is a chain terminating codon, and would only be expected to occur once if at all, in phase, in the translated parts of these viral RNAs, and indeed in monocistronic messengers in general. This might contribute to the relatively low frequency of (Gp)UpApGp in the viral RNAs. (Since there is no obvious cause for restriction on UpApGp occurring out of phase, the

overall frequency of UpApGp-terminated oligonucleotides should be roughly two thirds of strictly random expectation.) It is of interest that in polio RNA the frequency of the isomer, (Gp)ApUpGp is higher than expected. ApUpGp codes for initiating or internal methionine, but out of phase there is no obvious reason why this oligonucleotide should occur in other than random frequency.

In conclusion it might be suggested that continued analysis of RNAs for non-random trends in oligonucleotide frequencies, by these and related methods, might assist in tracing the relationship and possible evolution of viral RNAs from, or along with, host cell nucleic acid sequences.

#### Acknowledgements

We thank D.F. Summers and P. Fellner for gifts of  $^{32}\text{P}$  labelled poliovirus and EMCV RNA, and Professor J. Subak-Sharpe for discussion. This work was supported by a grant from the Medical Research Council. One of us (D.F.H.) is a recipient of a Science Research Council Studentship.

#### REFERENCES

1. Swartz, M.N., Trautner, T.A. and Kornberg, A. (1962). J. Biol. Chem. 237, 1961-1967.
2. Subak-Sharpe, J.H. (1967). British Medical Bulletin, 23, 161-168.
3. Subak-Sharpe, J.H., Burk, R.R., Crawford, L.V., Morrison, J.M., Hay, J. and Keir, H.M. (1966). Cold Spring Harbor Symposia on Quantitative Biology 31, 737-748.
4. Morrison, J.M., Keir, H.M., Subak-Sharpe, J.H. and Crawford, L.V. (1967). J. Gen. Virol. 1, 101-108.
5. Hay, J. and Subak-Sharpe, J.H. (1968). J. Gen. Virol. 2, 469-72.
6. Scholtissek, C. (1969). Biochim. Biophys. Acta 179, 389-397.
7. Scholtissek, C. and Rott, R. (1969). J. Gen Virol. 4, 565-570.

8. Subak-Sharpe, J.H., Elton, R.A. and Russell, G.J. (1974). Symposia for the Society of General Microbiology 24, 131-150.
9. Fraser, N.W., Maden, B.E.H. and Burdon, R.H. (1973). FEBS Letters 36, 257-260.
10. Burdon, R.H., Shenkin, A., Fraser, N.W., Smillie, E.J. and Douglas, J. T. (1974). FEBS Symposium, "The Cell Nucleus", in the press.
11. Porter, A., Carey, N. and Fellner, P. (1974). Nature 248, 675-678.
12. Brownlee, G.G. and Sanger, F. (1967). J. Mol. Biol. 23, 337-353.
13. Sanger, F., Brownlee, G.G. and Barrell, B.G. (1965). J. Mol. Biol. 14, 373-398.
14. Bishop, J.M., Summers, D.F. and Levintow, L. (1965) Proc. Nat. Acad. Sci. U.S. 54, 1273-1281.
15. Granboulan, N. and Girard, M. (1969). J. Virol 4, 475-479.
16. Tannock, G.A., Gibbs, A.J. and Cooper, P.D. (1970). Biochem. Biophys. Res. Commun. 38, 298-304.
17. Darnell, J.E., Jelinek, W.R. and Molloy, G.R. (1973). Science 181, 1215-1221.



# Instructions to authors

## Policy & organisation of the journal

Nucleic Acids Research publishes full papers in English which describe research on physical, chemical, biochemical or biological aspects of nucleic acids. The majority of papers acceptable for publication would fall within one or more of the following subject areas:

Physical and organic chemistry of nucleic acids.  
Their constituents and analogues.  
Nucleic acid biochemistry including enzymology.  
The role of nucleic acids in gene regulation and in virology.

Papers on properties of nucleotide coenzymes and simple purine and pyrimidine derivatives that have no relevance to the function and properties of the nucleic acids themselves will not normally be acceptable.

Papers should not normally exceed 20 pages in length. The typewritten manuscript will be reproduced photographically, which will eliminate the necessity to correct proofs and will result in fast publication (average 6-8 weeks). Each manuscript will be refereed by two scientists who have been instructed to try to recommend acceptance or rejection of a paper and to avoid suggesting minor alterations to content or style.

Preliminary communications or short communications which do not contain sufficient experimental documentation are not accepted. Submission of a paper implies that it reports unpublished work, that it is not under consideration for publication elsewhere and that, if accepted, it will not be published elsewhere in the same form either in English or in any other language without the consent of the publishers. Manuscripts (the original and two copies, between cardboards to avoid creasing) should be submitted thus:

*from the American continent to:*

Professor Dieter Soll  
Yale University, Box 1937 Yale Station  
New Haven, Connecticut 06520, USA

*from the rest of the world to:*

Professor A. S. Jones and Dr. R. T. Walker  
Department of Chemistry  
University of Birmingham  
PO Box 363, Birmingham B15 2TT, England

It is essential that manuscripts be prepared according to the instructions which follow. Manuscripts which do not conform to this specification may have to be returned for retyping.

## Preparation of manuscripts

Specially ruled typesetting sheets are available on request and the editors recommend their use by authors wherever possible.

It is the policy of the editors to allow a flexible style of presentation of papers but, to help us maintain a high standard of appearance and readability, please ensure that your typescript is in black (not grey), on good quality white paper and conforms to the following rules:

- (a) Type the title, author and address on a separate sheet of paper. This will be re-typed by the publishers to a standard format.
- (b) **Start page 1 with your abstract, which should not exceed 15 lines. Then start the main text of your paper. After filling a depth of 180mm (7in) including the abstract, start page 2.**

(c) **Continue your paper, filling each page to a depth of 240mm (9½in) except the last page, which may fall short.**

(d) The width of all copy (i.e. the longest lines) must not exceed 165mm (6½in).

(e) Each page should be numbered at the bottom in *light blue*. The dimensions quoted above include all diagrams/figures.

(f) Type the abstract and references at single line spacing.

(g) Type all main text at one-and-a-half line spacing (usually three 'clicks').

(h) Type headings to sections (e.g. ABSTRACT, INTRODUCTION etc) in underlined CAPITALS, aligned with the left-hand margin.

(i) **Indent the first line of each paragraph but do not leave any extra space between lines.**

(j) Align diagrams/figures and their legends and captions with the left-hand margin. Type the legends at single line spacing. Do not leave excessive space above or below diagrams/figures. Text and diagrams should normally be in sequence as appropriate to the sense of the paper. Legends should be brief and are not expected to make the figures meaningful without reference to the text. Diagrams can be submitted independently of the text, in which case they may then be photo-reduced by the publisher.

(k) Photographs are accepted when they are of high quality and are essential to the meaning of the paper. A charge will be made for publication of a colour plate.

(l) Footnotes, including abbreviations and changes of address, should be indicated with superscript figures and included in the references.

## References

The citation of journals<sup>1</sup> (abbreviated in the style of 'Chemical Abstracts'), books<sup>2</sup> and multi-author books<sup>3</sup> should conform with the following examples (single line spaced without indenting):

- 1 Razzell, W.E. and Khorana, H.G. (1958) *Biochim. Biophys. Acta* 28, 562-566
- 2 Davidson, J.N. (1969) *The Biochemistry of the Nucleic Acids*, 6th edn, pp. 177-178, Methuen, London
- 3 Burdon, R.H. (1971) in *Progress in Nucleic Acid Research and Molecular Biology*, Vol. II, pp. 33-79, Academic Press, New York

## Nomenclature

As far as possible, authors should follow the Recommendations of the IUPAC-IUB Commission on Biochemical Nomenclature, and particularly the abbreviations for nucleic acids, polynucleotides and their constituents (1971) *J. Mol. Biol.* 55, 299-305.

## Reprints

Upon acceptance a reprint price list and order form will accompany our notification. Thirty reprints will be supplied free of charge: additional reprints may be purchased in lots of 100 and, owing to the short production time for articles in this journal, an estimate of the additional number of reprints required should accompany each manuscript.



