RIBOSOMAL RIBONUCLEIC ACIDS OF XENOPUS LAEVIS

Jonathan M.W.Slack

Submitted for the degree of Doctor of Philosophy
University of Edinburgh
1974.

ACKNOWLEDGMENTS

The following helped me, materially and spiritually, during the course of the work described in this thesis:

Anne Baker
Rowena Clayton
Howard Cooke
Dennis Cremer
John Deag
Ron Fraser
Harvey Gordon
Phil Hamilton
Brian Higgins
Steve Hughes
Maria Luisa Lee
Sue Lewis
Peter Little
Ulrich Loening
Ron McKay
Alan Morrison
Ken Murray
Lynne Murray
Martin Rabstein
Caryl Robertson
Janice Robertson
Liz Rogers
Wolfgang Schuch
Dan Shapiro
Ed Southern
David Williams

Ulrich Loening was my supervisor and I was supported financially by the Ramsey Wright Scholarship of Edinburgh University.

"Hobbits liked books about things they already knew, set out fair and square with no contradictions"

Tolkien.

"Now, everybody is working away at 'projects' the outcome of which must be known in advance, since otherwise the inordinate financial investment could not be justified."

Chargaff.

SUMMARY

1. Various structural investigations on ribosomal RNA from Xenopus laevis are described. In the Introduction, the mechanism of synthesis of ribosomal RNA and the properties of the ribosomal genes as representing a multiple gene array are reviewed. The investigations described in the subsequent chapters centre on the structure of the 40s ribosomal precursor RNA, which is of especial interest as a well defined primary gene product; and on the consequences of the existence of multiple copies of the ribosomal genes for the structure of the ribosomal RNAs.

2. Xenopus kidney cells are shown to accumulate 40s RNA in the presence of cycloheximide. This is made the basis of a new procedure for preparation of quantities of the 40s RNA suitable for chemical studies.

3. The heterogeneity shown by 40s RNA on gel electrophoresis is shown to be reversed by denaturation, and to be exhibited also by 28s RNA. It is therefore believed to be caused by the existence of many metastable secondary structures which are fast to form and slow to interconvert.

4. The 5' termini of the 18, 28 and 40s RNAs are investigated by the method of alkaline hydrolysis. Although 5' triphosphates can be demonstrated in total RNA and in 5s RNA, the 40s RNA bears only a 5' monophosphate with the terminus pG. This suggests either that the 40s molecule is not a primary transcription product or that there exists some special feature of transcription of the ribosomal genes. The 5' termini of the 18 and 28s RNAs are both pU.

5. The 40s and 28s RNAs are shown to contain a sequence of three adjacent ribose methylations which is believed to be GmAmAmAp. This sequence is not at the 5' end. A similar methylated tetranucleotide is found in rat 28s RNA. This, together with the end groups and the 5' oligonucleotides recovered after labelling with polynucleotide kinase, disproves an earlier and much quoted report suggesting that the 28s segment lies at the extreme 5' end of the precursor molecule.

6. A quantitative approach is developed towards the oligonucleotide mixtures produced by ribonuclease digestion. Equations are derived to express the distribution of radioactivity on the isostich length, the distribution of compositional isomers on the isostich length and on intensity. It is shown what conclusions can be drawn about a sequence from a quantitative study of its digestion products.

7. The range of application of the equations is determined by computer simulations of isostich and compositional isomer separations. Experimental isostich separations and "fingerprints" are carried out for 18 and 28s RNAs and compared with the theoretical predictions. The results show that both sequences approximate to randomness.

8. The nature of "fingerprints" of transcripts of diverged repetitious DNA is considered, and a method is proposed for calculating the parameters of a sequence family from the fingerprint. The fingerprints of Xenopus ribosomal RNAs are discussed in the light of this model.

CONTENTS

Chapter 1

INTRODUCTION

The Structure of the Ribosomal Genes

This thesis is concerned with structural studies on the ribosomal RNAs of Xenopus laevis. There were two main reasons for this work. First, the ribosomal genes are the best known case of a multiple gene array, the properties of which pose several theoretical problems. Secondly, this remains the easiest system to work with for the study of the control of transcription, not least because of the existence of a well defined primary transcript, the ribosomal precursor RNA. Although the work to be described below is concerned exclusively with RNA, the inferences concern the genes, whose known properties are therefore carefully considered in this Introduction.

Xenopus laevis has been the species of choice for the study of the ribosomal genes and the recent work in this field may be said to have begun in 1958 with the discovery of the anucleolar mutant by Elsdale et.al. (1). The nucleolus is an intranuclear organelle associated with a chromosomal constriction called the nucleolar organiser, of which the normal somatic cells of Xenopus contain two. The mutation appeared to be a deletion of the nucleolar organiser. Homozygotes had no nucleoli and died at the early tadpole stage, heterozygotes were normal except that they possessed only one nucleolus per nucleus. In 1964, Brown and Gurdon showed that the anucleolate mutant lacked the ability to make ribosomal RNA (2). They proposed that the genes for both the 18s and the 28s RNA were located at the nucleolar organiser. It was already known at that time that there were many genes for ribosomal RNA because of saturation hybridiz-

ations carried out with bacteria (3) and pea seedlings (4), and this was confirmed by Birnstejil et.al.(5) for Xenopus. The currently accepted figure for the multiplicity of the ribosomal genes in this species is 600 (6).

That the anucleolate mutant had suffered a deletion of the ribosomal genes rather than a mutation in a regulatory gene was shown by Birnstejil's demonstration that the genes formed a heavy satellite (p=1.723) on CsCl gradients (5). The DNA from wild type, heterozygous and homozygous recessive individuals showed the satellite in a ratio of 2:1:0. The formation of a satellite in DNA of high molecular weight showed that the genes must be clustered, perhaps all together.

In 1968, the groups of Brown and Birnstejil reported essentially similar experiments which demonstrated the existence of "spacer" DNA in between the ribosomal coding regions (7,8). First, although 18s and 28s RNA hybridized in the same position to a CsCl gradient of large pieces of DNA (mol.wt. 4×10^6), these positions were separated when the DNA was sheared to a small size (mol.wt. 0.3×10^6). Both move to a less dense position and the genes for 18s RNA move slightly further. Therefore the 18s and 28s coding regions must be interspersed, and be separated by a spacer of very high CG content. Similar conclusions were drawn from the second experiment in which DNA of different sizes was pre-hybridized to a cold rRNA, fractionated on CsCl and challenged with radioactive 18s and 28s RNA. In this experiment the banding position of the pre-hybrid depended on its RNA: DNA ratio.

That the 18s and 28s coding regions were strictly alternating was strongly suggested by the discovery of the ribosomal precursor RNA in mammalian cells (9,10,11).

This was shown to contain both sequences by hybridization competition (12). Such a "polycistronic" precursor was discovered in Xenopus in 1969 (13,14). In the same year Miller and Beatty(15) published electron micrographs of DNA from the extrachromosomal nucleoli of Triturus viridescens. These showed an axial fibre with "matrix units" of about two microns alternating with matrix free units. The matrix units consisted of polarized fibres attached to the axial fibre, which were longer at the distal than at the proximal end. Experimants involving nuclease digestion and labelling with radioactive RNA precursors indicated that the axial fibre was DNA and that the fibres of the matrix unit were growing chains of RNA. A "blister map" of ribosomal DNA was constructed in 1971 by Wensink and Brown (16). They employed electron microscopy of DNA partially denatured in alkali and found the repeating unit to be 5.4 microns in length.

The arrangement of the 18 and 28s segments in the repeating unit has recently been established by electron microscopy of partially denatured precursor RNAs (I.Dawid personal communication). In addition, the 5.8s RNA coding region has been localised by shearing and hybridization experiments (J.Speirs, personal communication). The structure of the repeating unit now seems to be as shown in Fig. 1.1.

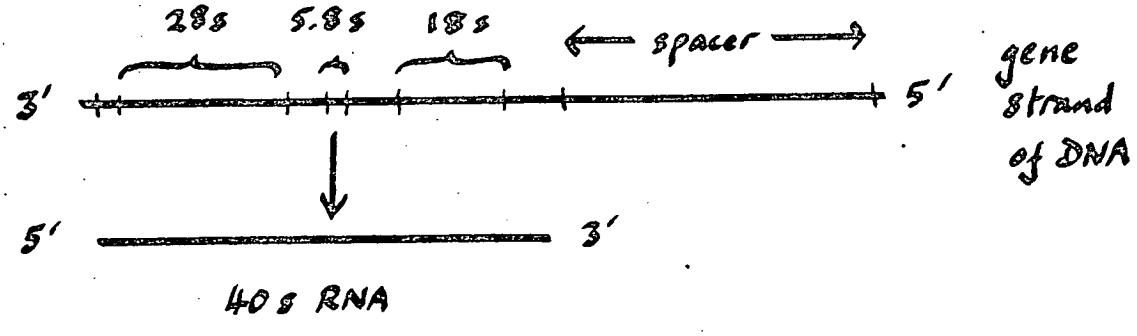


Fig.1.1 Arrangement and polarity of the ribosomal repeating unit.

Multiple Gene Arrays

The ribosomal genes form one example of several multiple gene arrays now known to be present in higher organisms. Other molecules coded for by multiple DNA sequences include the histones (17), transfer RNA (6) and 5s RNA (18). The genes for immunoglobulin V regions seem to exist in the germ line as a large number of related sequences (19) and in addition the higher organism genome contains a large fraction of kinetically defined "reiterated DNA" (20) of which at least some is transcribed (21), so it is quite possible that additional multiple gene arrays remain to be discovered.

There are three classes of theoretical problem which are associated with the multiple gene arrays, and they are not usually clearly distinguished:

i) How are the multiple copies formed in the first place?

ii) How do they spread through an interbreeding population ?

iii) What is their mode of evolution after they have become established ?

Before discussing these problems we shall briefly review the properties of the better known cases, starting with the ribosomal genes.

In Xenopus laevis, the ribosomal repeating unit is present in about 600 copies per haploid genome (in other species the numbers vary between 160 for yeast and 14000 for the Chinese cabbage (22)). These copies are clustered at the nucleolar organiser. There is no melting point depression after denaturation and reassociation of the rDNA of either a single individual or of a group of individuals in the same species (23), which suggests that the multiple copies are very similar or identical. The group

of D.D.Brown have studied the ribosomal genes of a related species, Xenopus mulleri (24). These are organized in the same way as the genes of X.laevis, and the rRNA of the two species, hence also the coding regions of the genes, are believed to be identical. However, the spacer regions (including the transcribed spacer which will be referred to hereinafter as "excess") are very different by base composition, by hybridization to heterologous cRNA and by comparison of "blister maps". So it would seem that the ancestor to the two species had the repeating unit arrangement and that after the reproductive isolation the spacer regions have rapidly diverged in sequence while the coding regions have remained conserved. Each change has been fixed both vertically (in the population) and horizontally (in the tandem repeats).

The 5s RNA genes of Xenopus present a similar picture with one important difference. The genes from X.laevis were isolated by Brown et.al. (18) by a procedure involving four equilibrium gradients. The genes have a much lower CG content than would be predicted from the 5s RNA sequence, and a study of the blister map suggested that the spacer was six times the length of the coding region. The 5s genes from X.mulleri were described two years later (25). From saturation hybridization and CsCl banding of the hybrid, the spacer was found to be much longer than that of X.laevis, about seventeen times the length of the coding region. Cross hybridization with cRNA showed that the spacers were greatly different in sequence while the coding regions were conserved. So far the situation seems to be the same as that of the ribosomal genes except for the different lengths of the spacers. But the 5s genes differ in that they do not appear to be identical by the criterion of melting point depression, either within the species or within the individual organism.

Some work on the histone genes has been reported by Birnstejl and co-workers (26,27). Three histone messenger RNAs were purified from early embryos of the sea urchin Psammechinus milaris. All hybridized to the same position on the dense side of a CsCl gradient of the DNA, and all hybridized to sequences with a multiplicity of about 1200 under conditions of DNA excess. So it would seem that these genes also are present in clustered multiple copies. Because these genes are very highly conserved in evolution, it proved possible to perform an in situ hybridization on Drosophila salivary gland chromosomes, and to show that all the messages hybridized to a short segment of chromosome 2.

The histone genes have not yet been isolated and so there is no melting point depression data on the question of divergence between the multiple copies. However, the RNA-DNA hybrids are of high thermal stability which indicates considerable similarity and certainly more than would be expected if the genes were diverged to the extent permitted by the requirements of the protein sequences. A T_m depression of about 5^o was found for interspecific RNA-DNA hybrids between Psammechinus milaris and Paracentrotus lividus. If this is taken to represent a sequence divergence of about 3%, then it is about one quarter of that of the bulk DNA, which is the expected value if there were selection for exact conservation of the protein sequences.

There are no known genetic markers for ribosomal or 5s genes, but one is known for a histone. This is position 15 in histone V from the chicken (28). Histone V is a protein which occurs only in the nucleated erythrocytes of vertebrates, and if its genes form a multiple cluster like those for the other histones then the existence of a Mendelian marker strongly suggests that the copies are identical. In this case

they must be subject to the same processes of horizontal fixation as the ribosomal genes.

A beginning has been made to the study of the transfer RNA genes in Xenopus (29,6,30). Each tRNA gene is reiterated about 200 fold, but the multiplicities vary, even between different isoacceptor tRNAs. Prehybridization of excess, unfractionated tRNA to DNA leads to a small but definite density increase for each of the various genes and so it is presumed that each tRNA species has clustered genes with quite long (0.5×10^6) spacer regions. Probably the spacers for the different clusters are different but it is not yet possible to say whether there is a single cluster for each tRNA gene or whether the tandem repeats are identical.

Divergence between clusters

The implication behind our consideration of the four types of gene together is that rRNA, 5s RNA, tRNA and histone genes all share a common set of characteristics, namely that they are composed of identical or very similar repeating units which are clustered, each repeating unit containing a coding region and a spacer region.

The main problem about this conclusion is the heterogeneity of the Xenopus 5s genes which was alluded to above. This heterogeneity is also evident from the primary structure of the 5s RNA (31,32). To explain this we need not invoke differences between individual copies of the 5s genes but only differences between isolated clusters. It has been shown by in situ hybridization of cRNA from isolated 5s DNA that the 20,000 gene copies are divided into many clusters present on most of the chromosomes, in the telomere regions of the long arms (33). The primary sequence

data in fact strongly suggests a situation in which there are certain differences between internally homogeneous clusters. 5s RNA from kidney cells was found to have a unique sequence, while that from the ovary contained certain partial substitutions at various positions along the molecule. So perhaps only one cluster is transcribed in the former tissue and several clusters in the latter.

Xenopus has only one cluster of ribosomal genes per haploid chromosome set. But mammalian species may have several nucleolar organisers per haploid set, and so we might here expect to find heterogeneity in ribosomal RNA caused by differences between the gene clusters. There are in fact reports of base composition and sequence differences between mouse rRNAs from normal tissues and tumours (34,35), low thermal stability of rRNA-DNA hybrids in the rabbit (36), and complex renaturation kinetics in the rat (37). However in Man, an organism with six nucleolar organisers per chromosome set (38), there seems to be little or no divergence in the rDNA (39), so this question must be regarded as unsettled.

Possible solutions to problems

It is not known how the multiple copies first arose or how they spread through the population. The chances are that these events occurred in remote antiquity since the multiple genes are probably ubiquitous and the known ones code for molecules which are present in all living organisms and required in large quantities. The satellite DNAs (40) and the immunoglobulin V genes (19) present a contrast to this. They show great differences of sequence family pattern between species and considerable divergence between neighbouring sequences within a cluster. This suggests that they are subject to ongoing evolutionary processes of

multiplication and divergence.

When it comes to the question of sequence conservation within multiple gene clusters, several theories have been propounded. These will be listed and then it will be explained why they are all inadequate.

i) Callan's "master-slave" hypothesis (41) which was originally stated in connection with lampbrush chromosome loops. Here, in each individual, all the copies are corrected in accordance with a master copy.

ii) Brown's "expansion-contraction" theory. The number of ribosomal genes is drastically reduced at intervals in evolution and is then restored by amplification of the few remaining copies.

iii) Smith's unequal crossing over theory (42). Here unequal crossovers between identical arrays in sister chromatids causes a small fluctuation in gene number with cell generation which eventually leads to homogeneity in a single tandemly repeating set.

iv) An unpublished theory of Birnstiel's required that a point mutation inhibited recombination over a long stretch of the array, and therefore isolated itself from the gene pool. As soon as it became linked to a lethal mutation it became eliminated from the population.

There is no direct evidence in favour of the master-slave theory and recently some good evidence has been obtained against it (D.D.Brown, unpublished). When X.laevis-X.mulleri hybrids are backcrossed to one of the parent strains, all the offspring are found to contain spacer DNA from both of the species specific types. This proves both that recombination occurs in the ribosomal gene array and that there is no mechanism for correction of the abnormal copies at the level of the individual.

Birnstiel's idea is an ingenious one, but the

genetic evidence that recombination can occur between very close point mutations makes the main premise rather unlikely.

Theories ii. and iii. are essentially similar, the former involving more drastic fluctuations of gene number at longer intervals of time. The main experimental evidence is that provided by Ritossa(e.g.43) who has shown that strains of Drosophila of a strong bobbed phenotype will recover over several generations, and that this recovery is associated with the restoration of the normal complement of ribosomal DNA in the genome. However, in the opinion of the author, these theories do not explain anything at the moment because they are not posed in population terms. The computer simulations described by Smith (42) do not represent a biological population but only a single lineage of chromosomes in a population of somatic cells. Homogeneity among tandem genes in a population of organisms would be achieved only by the normal mechanism of genetic drift which depends on the small sampling errors in those gametes to be represented in the next generation. Unequal crossing over might perhaps have a role in the generation of a pseudo-allelism on which genetic drift could operate.

The other notable evolutionary feature of the ribosomal genes is that the spacer region has undergone much faster interspecific divergence than the coding region. This, however, is not so mysterious as the conservation problem because it can be explained if most of the mutations in the coding sequences are deleterious while those in the spacer are neutral or nearly so.

Other genes

The aforementioned work on multiple gene arrays raised the suspicion that many or all genes might be present in multiple copies, particularly in view of the large amount of DNA per cell in higher organisms and the fact that the bands in polytene chromosomes and the loops in lampbrush chromosomes behave to some extent like genes although they are much too big. Gene multiplicities can now be measured by the hybridization of isolated mRNAs, or better their DNA complements, to a vast excess of DNA. Experiments of this type with the mRNA for haemoglobin (44), silk fibroin (45) and ovalbumin (46) indicates that only a single gene copy exists per chromosome set.

Synthesis of Ribosomal RNA

This subject has been reviewed repeatedly in recent years (47,48,49,50,51,52). The most detailed knowledge exists for the HeLa cell and it is now clear that the processing scheme is not necessarily the same in other eukaryotes. However, there are a number of general features for which the evidence is listed below:

i) Ribosomal RNA is made in the nucleolus.

ii) The first product of transcription is a high molecular weight precursor RNA.

iii) Transcription is inhibited by low concentrations of actinomycin D.

iv) rRNA is made by a different RNA polymerase from mRNA.

v) Various modifications are made to the precursor molecule shortly after transcription; these include several ribose methylations, a few base methylations, and conversions of uridine to pseudouridine.

vi) The precursor contains one and only one of each of 18s and 28s RNA sequences.
vii) The precursor is invested with proteins and cleaved at specific sites, eventually forming 18 and 28s RNAs.
viii) The ribosomal subunits are assembled in the nucleolus and then transported to the cytoplasm.

Evidence

i) The uptake of ^{3}H-uridine into nucleoli is blocked by low actinomycin (53), there is no rRNA synthesis in the anucleolate Xenopus (2), and all the various rRNAs are made in isolated nucleoli (54).
ii) A pulse label of ^{3}H-uridine goes into RNA of about 45s which in HeLa cells appears in about 5 minutes and reaches a maximum specific activity in about 15 minutes. This material has a similar base composition to rRNA and the label may be chased into 18+28s RNA in the presence of low actinomycin (55).
iv) RNA polymerases may be separated on DEAE Sephadex. In all the higher organisms studied the first component is nucleolar, functions in low salt and is insensitive to α-amanitin, while the second component is nucleoplasmic, functions in high salt and is sensitive to α-amanitin. Polymerase I is presumed to be responsible for rRNA synthesis (56).
v) rRNA synthesis can be followed using (^{3}H-methyl) methionine (56) which characteristic was used to estimate the transcription time of the 45s molecule as 2.5 minutes. The methylations in HeLa rRNA were first analysed by Wagner et.al.(58).
vi) The presence of the 18s and 28s sequences in the 45s RNA was shown by hybridization competition (12).
vii & viii) A series of ribosomal precursor particles has been found in the nucleolus (59) the first of which contains both 45s and 5s RNAs, the 5s RNA coming from

outside the nucleolus. Processing and transcription are inhibited in concert by cycloheximide (60). The specificity of the cleavage reactions is indicated by the homogeneous terminal sequences reported for ribosomal RNAs (49).

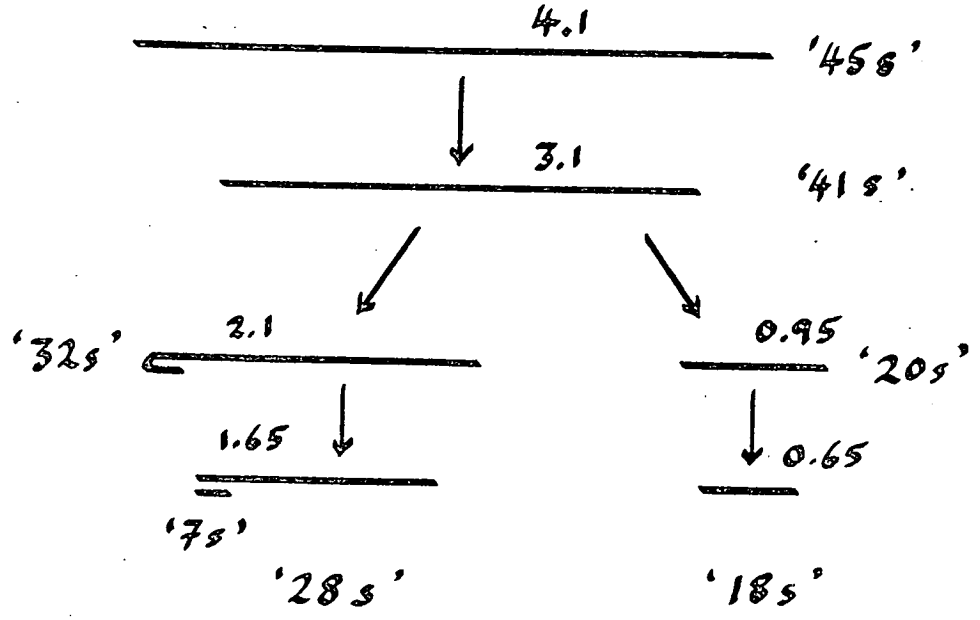


Fig.1.2. Processing scheme for HeLa rRNA proposed by Weinberg et.al.

The detailed analysis of the processing scheme in HeLa cells came when gel electrophoresis was introduced as a fractionation technique for RNA (61). This had a higher resolution than sucrose gradients and enabled the measurement of molecular weights by the semi-log relationship with mobility (62). Weinberg et.al. (63,64) found nucleolar RNAs of molecular weights 4.1, 3.1, 2.1, 1.4 and 0.95 million daltons, all of which were shown to be precursors of rRNA by chase kinetics in actinomycin D. Infection with poliovirus in the presence of 2 mM guanidine caused the accumulation of

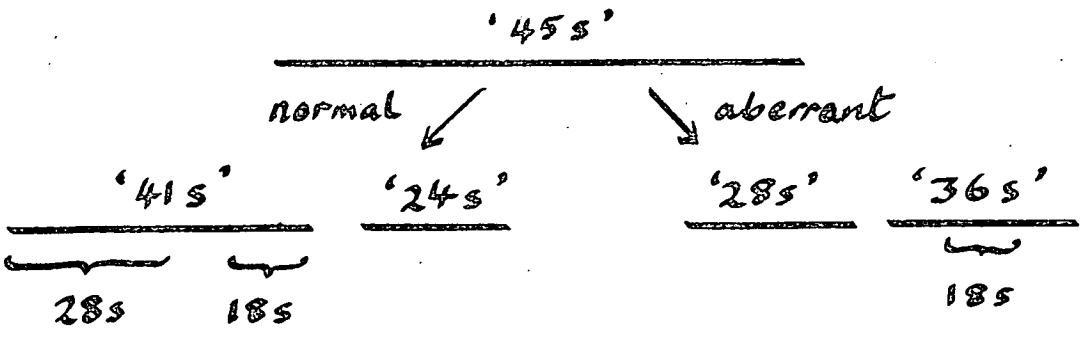


Fig.1.3. Formation of 24s and 36s RNAs by an incorrect cleavage order.

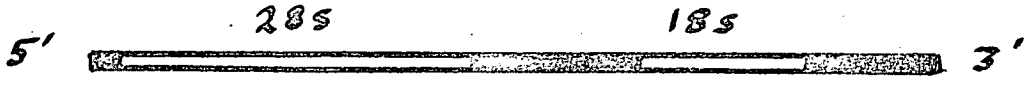


Fig.1.4. Structure of 45s RNA determined by electron microscopy.

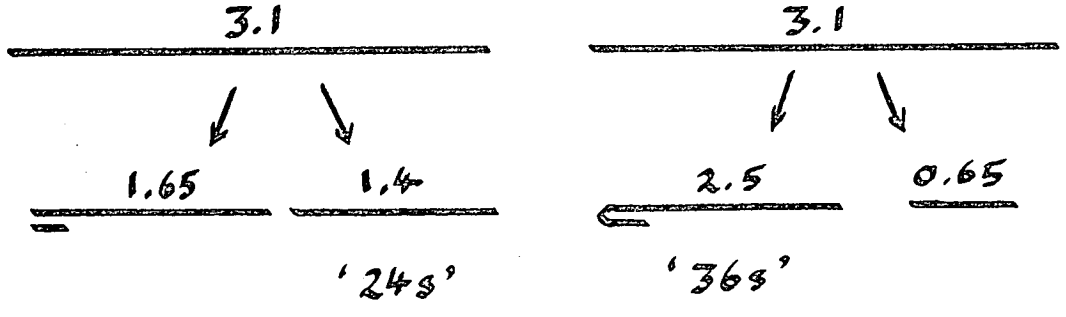


Fig.1.5. Formation of 24s and 36s RNAs according to Wellauer and Dawid.

the 41, 28, 20 and 18s RNAs so that their ^{32}P/methyl label ratios could be accurately measured. On the basis of the molecular weights and the P/Me ratios the processing scheme was proposed which is shown in Fig.1.2. The 2.5 and 1.4 species were presumed to arise from premature cleavage of the 41s (Fig.1.3).

In 1973, Wellauer and Dawid (65) examined all the HeLa nucleolar RNAs by electron microscopy. After partial denaturation, the 28s and 18s sequences are recognisable by distinct secondary structure features and the polarity was assigned by digestion with a 3' exonuclease. The 45s precursor had the structure shown in Fig.1.4, and a different origin is proposed for the 36s and 24s RNAs. The former is supposed to arise from the 3' end of the 45s RNA and the 24s to be the normal excess fragment from the same end (Fig.1.5).

Ribosomal RNA processing has now been studied in several animals and plants (eg.66) and it seems likely that in no case will the scheme turn out to be exactly the same as the HeLa cell. In particular, organisms other than birds and mammals seem to have much smaller primary precursors (67). In this account we shall consider only the case of Xenopus because of the wealth of knowledge about the structure of the genes in this species. A high molecular weight rRNA was first found by Gall (68). It was shown to be a ribosomal precursor by actinomycin chase kinetics and labelling with (^{3}H-methyl) methionine by Landesmann and Gross (13). Using gel electrophoresis, Loening et.al. (14) and Rogers and Klein (69) measured the molecular weight as 2.6 million daltons and found two precursors of 28s RNA (mol.wts. about 1.78 and 1.6 million) but no immediate precursor to 18s RNA. Electron microscopic analysis by I.Dawid (personal communication) reveals in addition a 38s RNA which contains both the 28s and

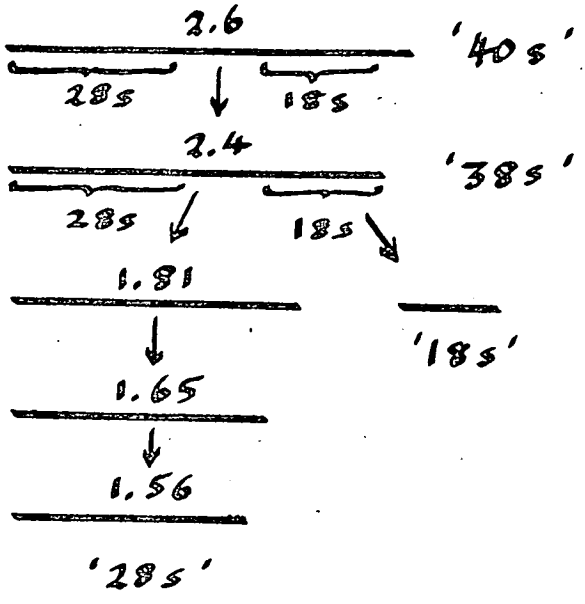


Fig.1.6. Processing scheme for Xenopus rRNA proposed by Dawid.

18s sequences (see Fig.1.6).

Quite a battery of enzymes must be involved in the entire process of ribosome formation, quite apart from the ribosomal proteins. In particular we can expect to find some specific methylases, exo- and endonucleases. The methylation of HeLa rRNA has been carefully studied by Maden et.al. (70,71) and it is now clear that all the sites are quantitatively methylated. Vaughan et.al. (72) showed that processing was arrested at the 32s stage in methionine starved cells, and that this 32s RNA was undermethylated. Perry and Kelley (73) have described a 3' exonuclease from L cell nucleoli which is perhaps responsible for the exonucleolytic steps of processing and which is inactive against 2'-O-methylations. There are also some preliminary results which suggest the presence of specific nucleolar endonucleases (74,75,76). These are expected by

analogy between RNA processing and microbial restriction-modification systems (77), but it may well prove to be that the specificity resides as much in the secondary structure of the RNA or in RNA-protein interactions as in the cleavage enzymes.

Relevance of the Thesis

Many of the results mentioned above were not available when this work was commenced in 1971. In particular the arrangement of the 18 and 28s sequences on the ribosomal precursor was unknown, and the electron microscopic comparison of the ribosomal genes of X.laevis and X.mulleri had not yet been done. The latter work indicates that all the tandemly arranged ribosomal genes are the same to the level of resolution provided by electron microscopy, but in 1971 various pieces of evidence seemed to point to heterogeneity, especially of the precursor RNA.

It had been demonstrated in this laboratory that the ribosomal precursor from Xenopus and from Schizosaccharomyces pombe showed heterogeneous behaviour on gels. The RNA from a particular position in the peak would remain at this position on rerunning (78). Also ribosomal precursors of different lengths had been found in different tissues of Phaseolus aureus (79), and it had been shown that the gel mobility of the mouse precursor depended on the length of the radioactive pulse used to label it (80).

So, it seemed important to characterize the heterogeneity of the precursor RNA. Xenopus was selected as the experimental organism for several reasons:

i) It was an organism in which the heterogeneity had

been demonstrated.
ii) It has only one cluster of ribosomal genes per chromosome set, which reduces the number of possible causes for heterogeneity.
iii) The kidney cells grown in the laboratory contained relatively large amounts of 40s RNA when compared to other sources.

The main method of investigation adopted was that of 5' end determination. This should reveal a heterogeneity of length (polydispersity) because all four nucleotides would be found at the 5' end instead of just one. It would not distinguish between the cases where heterogeneity is determined by secondary structure effects and by a low level of sequence divergence (the latter is to be expected in several of the theories of multiple gene evolution discussed above). But a choice between these possibilities can be made by means of various denaturation experiments which are described in chapter 3.

The 5' end determinations, which are described in chapters 4 and 5, were also employed to answer questions about transcription and about the arrangement of the ribosomal coding regions. The 40s RNA is the first identifiable gene product and if it is a primary transcript it should bear a 5' triphosphate, so some experiments were carried out to look for this. Secondly, there was the question of whether either of the 18 and 28s sequences lay at the extreme 5'end of the precursor. At this time there was one report of a common methylated sequence from the 5' ends of the 28s and 45s RNAs in the rat (81), although most workers considered that the 18s sequence lay at the proximal end of the precursor molecule (23,82,83). Some preliminary experiments were performed on 3' end determination, but these were abandoned when it became

clear that the information gained would not justify the expenditure of time and effort.

The second part of this thesis is on a slightly different aspect of the multiple gene problem. The end group project had made the author familiar with many of the techniques used in the sequencing of RNA. He was struck by the qualitative nature of the field and wondered whether a more quantitative approach would enable the properties of the nucleic acids under analysis to be deduced from the statistics of the components of the digest. A statistical theory of digests was duly constructed and compared with computer simulations of unique and multiple sequence fingerprints.

Isostich separations and fingerprints were carried out for the high molecular weight RNAs with which the author had been working, and these were interpreted in the light of the theory.

Chapter 2

THE CYCLOHEXIMIDE EFFECT AND ITS APPLICATION TO THE PREPARATION OF 40s RNA

The vast majority of studies on the synthesis and processing of ribosomal RNA have involved the fractionation of whole cell, nuclear or nucleolar RNA as an analytical rather than a preparative method. As far as the preparation of the primary precursor RNA (40s in Xenopus) is concerned, gel electrophoresis has not the capacity and sucrose gradients have not the resolution. For those experiments which have in the past involved manipulable quantities of ribosomal precursor RNA it has been prepared by several cycles of sucrose gradient sedimentation (eg.23,73). This has the disadvantage that it is time consuming, exposes the RNA to considerable risk of degradation, and separates only by molecular weight.

In order to improve on this, three questions were considered:

i) How to increase the amount of starting material.

ii) Whether the amount of 40s RNA per cell could be increased.

iii) Which techniques to use, and in what combination, to obtain a rapid and efficient fractionation.

The final protocol enables the preparation of about 50 micrograms of 40s RNA per batch. It is presented diagrammatically in Fig.1 and is described below. After the description, some of the work which led up to it will be outlined. The description covers all the RNA preparations mentioned in this thesis; however a particular preparation may have been scaled down and for a given experiment the RNA species were only purified

Stage	Operation		Purpose
1	Cell growth. Cells given 1 µg/ml cycloheximide 2 hours before extraction.		Doubles 40s RNA pool.
2.	Cells dissolved in "TNT".		Cells form clear solution.
	Phenol extraction three times.		Removes protein.
	Ethanol precipitation.		Removes soluble fraction.
	Reprecipitation.		Removes phenol.
	DNAse treatment Ethanol precipitation.		Removes DNA.
3.	Run on 5-20% sucrose gradient in B XIV zonal rotor, 3 hours 10 mins, 45K. Ethanol precipitation.		Separates soluble, 18s, 28s and 40s RNA. Removes carbohydrate.
	Soluble peak	40s peak	
4.		Passed through oligo-dT cellulose. EtOH precipitation	Removes poly A-containing HnRNA.
5. 6.	Run on 7.5% polyacrylamide gel. RNA recovered by electrophoresis.	Run on 2.4% polyacrylamide gel. RNA recovered by electrophoresis.	Removes contaminating RNA species.

Fig.2.1. Protocol for preparation of ribosomal RNA from Xenopus kidney cells

as far as was required. For sterile preparations, all the glassware was autoclaved or washed with diethyl pyrocarbonate (DEP). Gloves were worn at all stages and solutions were Millipore filtered or treated with DEP at 60° for one hour.

Description of the Protocol

Stage 1

The Xenopus cells were grown in Winchester bottles on a Matburn rolling apparatus. The growth medium was:

MEM (Glasgow modification)	xl.
Foetal Bovine Serum	10%
Sodium bicarbonate	0.35%
Glutamine	2 mM
Penicillin+Streptomycin	50 units/ml.
Non-essential amino acids	xl.

(media from Biocult Laboratories, Glasgow)

This mixture was diluted by 16% with distilled water. 100-200 mls of medium were used per Winchester and the cells were grown for about a week. For preparation of ribosomal precursor RNA, 2 μg/ml of cycloheximide were given two hours before extraction.

Various radioactive labels were used in the course of this work. Labelling with ^{32}P was carried out in phosphate free medium containing about 50 μCi/ml of isotope. For pulse labels (1-3 hours) the medium contained 2% serum, and for long labels (24 hours or more) it contained 5% serum. Labelling with ^{3}H-uridine was carried out in the standard medium. The concentrated medium was diluted by 16% with the isotope instead of

with water and so this gave a final concentration of 0.16 mCi/ml. For methyl labelling, (^{3}H-methyl) methionine was used at 20 μCi/ml in methionine free medium with 5% serum. This medium also contained 20 μM adenosine, 20 μM guanosine and 20 mM sodium formate which is reported to suppress the labelling of purine rings (70).

Stage 2

Extraction of RNA was carried out by a modification of the method of Parish and Kirby (84). The cell sheet was washed in Dulbecco's phosphate buffered saline and dissolved in "TNT" which consists of:

Sodium triisopropylnaphthalenesulphonate	2%
Sodium chloride	1%
Tris chloride pH 7.4	50 mM

About 1-2 mls were used per 20 cm^2 of cells, and if there was a large amount of cells the mixture was homogenized. This solution was extracted twice with a phenol mixture:

Redistilled phenol	500 gms
Redistilled m-cresol	70 gms
8-hydroxyquinoline	0.5 gms

+ excess distilled water to saturate.

0.1 volumes of 3 M NaCl were added before the second extraction. A third phenol extraction was performed using a 1:1 mixture of the above "phenol" and chloroform.

The nucleic acids were precipitated from the aqueous phase with two volumes of ethanol and left overnight in the cold room. This precipitate was redissolved in 0.15 M sodium acetate pH 6, 0.5% SDS and reprecipitated with ethanol.

The second precipitate was washed in 80% ethanol, 0.1 M NaCl to remove SDS and was then dissolved in a small volume of MES buffer:

2-(N-morpholino)ethane sulphonic acid	50 mM	pH 7.0
Sodium hydroxide	37 mM	
Magnesium acetate	2.5 mM	

(sterilized by Millipore filtration)

20 µg/ml of DNAse I was then added and the mixture shaken at 0° for one minute or for fifteen minutes depending on the extent of DNA breakdown required. Then one volume of 0.15 M sodium acetate pH 6, 0.5% SDS was added and four volumes of ethanol to precipitate the RNA.

At this stage the preparation is referred to as "total RNA" although it also contains quite a lot of carbohydrate. There are several experiments in this thesis which use such a preparation.

Stage 3

The total RNA was dissolved in 10 mls of 2.5% sucrose gradient solution:

Tris chloride pH 7.4	50 mM
Sodium chloride	150 mM
EDTA	1 mM
SDS	0.5%

It was applied on top of a 5-20% gradient in the MSE B XIV zonal rotor and overlain with 100 mls of gradient buffer. The rotor was spun at 45K for 3 hours 10 minutes at 20°C. Ten ml fractions were collected and the effluent was monitored at 254 nm with an ISCO recording spectrophotometer. The fractions containing

the required RNA species were pooled and precipitated with two volumes of ethanol. Because of the large volume of the zonal gradient, the samples were rather dilute, especially the 40s RNA. So they were left at -20^{o} overnight to precipitate and were pelleted by spinning at 10K for 30 minutes in the High Speed 18 centrifuge. At this stage the 18 and 28s RNAs were pure except for a small amount of carbohydrate, but the 40s RNA required further purification.

Stage 4

Oligo-dT cellulose was equilibrated in "binding buffer" for a few hours:

Tris chloride pH 7.4	10 mM
EDTA	1 mM
SDS	1%
Sodium chloride	0.4 M

A column of 0.2 ml volume was poured into a 1 ml syringe. The sample was dissolved in 0.5 mls of binding buffer and passed through slowly. This removes poly A containing RNA species. The eluate was precipitated with two volumes of ethanol.

Stage 5

A large 2.4% polyacrylamide gel was made as described in ref. (61). The sample was dissolved in less than 100 µl of ½E buffer + 12% sucrose and applied to the top of the gel. E buffer is:

Tris hydroxide	36 mM	pH 7.6
Sodium dihydrogen phosphate	30 mM	
EDTA	1 mM	
SDS	0.2%	

Electrophoresis was continued for four hours at 50 V. The gel was washed in E buffer for about one hour and scanned at 260 nm in a Joyce Loebl gel scanner. If the sample was labelled with ^{32}P it was frozen in a metal trough with CO_2 snow and sliced to 1 mm slices in a Mickle slicer. The slices were put into scintillation vials and counted by Cerenkov radiation in the Packard 3320 scintillation counter (Gain 55%, full window).

The RNA was recovered from the slices of interest as follows. The slices were packed into a sawn off Pasteur pipette with muslin plugs at top and bottom. A dialysis bag was attached to the bottom of the pipette and the whole inserted into an electrophoresis tank. The current was left on until all the radioactivity had entered the bag, and then the contents were precipitated with two volumes of ethanol. This product is pure 40s RNA except for a little non-poly A containing HnRNA.

Stage 6

This stage was for the preparation of 5s RNA. The procedure was identical to that of stage 5 except that the gel was 7.5% acrylamide and was run for 2 hours. The sample was the precipitated "soluble RNA" from the zonal rotor.

Optimization of the Protocol

Amount of Starting Material

Xenopus cells grown *in vitro* contain perhaps 0.5% wet weight of RNA, of which no more than 1% is 40s RNA. So to prepare 1 µg of 40s RNA we need at least

20 mgs of cells. Since the cells grow in monolayer it was thought that tissues from the animal might be a more suitable starting material.

RNA was accordingly prepared from the livers of mature and very young toads. The extraction procedure was the same as for the cells except that the tissue was mechanically homogenized in the TNT buffer and the solution was centrifuged at 30K for 30 minutes to remove glycogen before the second ethanol precipitation. Unfortunatly there was very little 40s RNA in these preparations. None was detactable in the adult liver RNA and only a suspicion in the young liver RNA (Fig 2.2).

So it was decided to use cells after all and to grow them in a larger bulk than was possible in the plastic dishes then in use. A Matburn rolling apparatus was modified by rearranging the wheels and fitting several plastic belts so that two Winchesters could be rotated. These were used to grow the cells, the total surface area being equal to about twenty of the plastic dishes. The apparatus could yield up to one gram of cells.

Amount of 40s RNA per Cell

Willems et.al. (60) showed that in HeLa cells the processing of the primary precursor depended on continuing protein synthesis. It therefore seemed possible that an inhibitor of protein synthesis such as cycloheximide might cause an accumulation of 40s RNA, although this does not happen in HeLa cells.

The effect of cycloheximide on rRNA synthesis in Xenopus was investigated as follows. First, the total RNA was examined on polyacrylamide gels after a variety of treatments. It was felt to be especially important

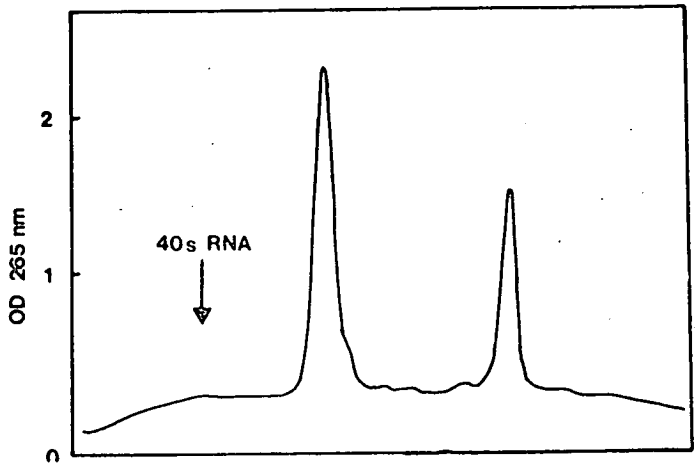


Fig. 2.2 Polyacrylamide gel of total RNA prepared from livers of young toads.

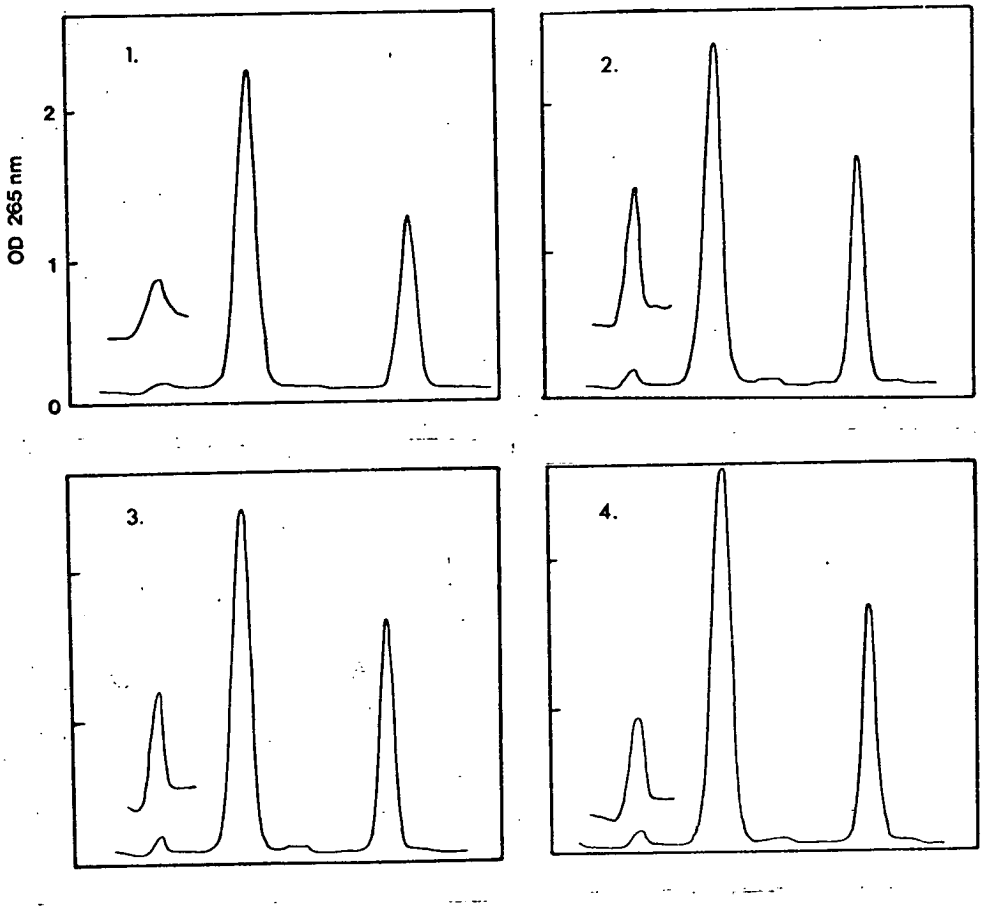


Fig. 2.3 Accumulation of 40s RNA caused by various concentrations of cycloheximide. 1) none 2) 2 µg/ml 3) 10 µg/ml 4) 20 µg/ml.

to examine the optical density rather than the radioactive profiles because the relative sizes of radioactive peaks depend on specific activity as well as on relative abundance. The results are shown in Figs. 2.3 and 2.4. It can be seen that cycloheximide does cause an accumulation of material at the 40s position, the maximum accumulation occurring after two hours in 2 ug/ml of inhibitor. It was found that this concentration of cycloheximide inhibited protein synthesis by 93% (Fig 2.5).

In order to prove that this accumulation was in fact the ribosomal precursor RNA, a pulse-chase experiment

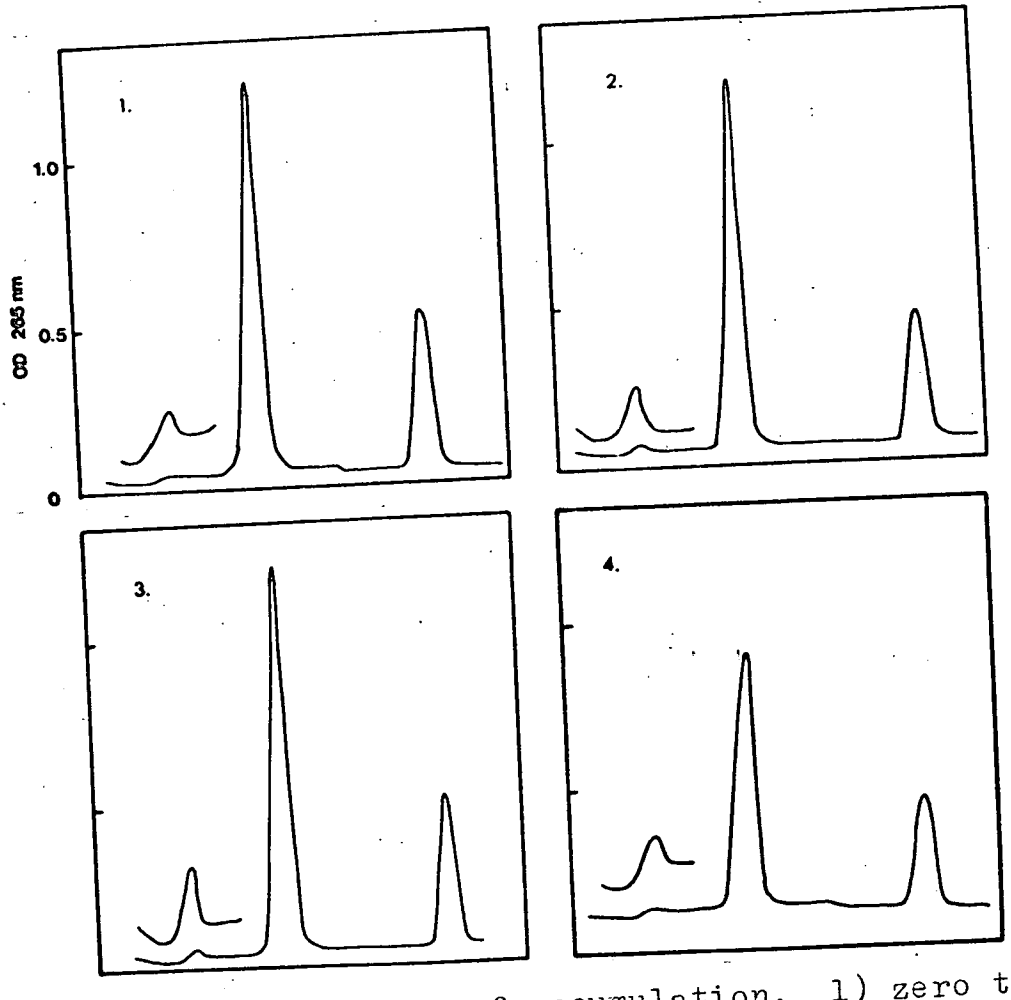


Fig. 2.4 Time course of accumulation. 1) zero time 2) 1 hour 3) 2 hours 4) 3 hours. The concentration of cycloheximide is 2 ug/ml.

was designed (Figs. 2.6 & 2.7). There were five dishes of cells and all received a 30' pulse of $^{32}P_i$. Dishes 4 and 5 show the state of labelling after a 1½ hour chase with and without cycloheximide. Dishes 2 and 3 show the state of labelling after a 5 hour chase. It is clear that the inhibitor greatly reduced the rate of processing of 40s RNA. In dish 1 the cycloheximide was removed after 1½ hours and it can be seen that processing has restarted, so the effect must be reversible.

Stage 2

RNA was always prepared from whole cells rather than from nuclei or nucleoli. There were two reasons for this: first the cytoplasmic RNAs were often wanted as well as the nucleolar ones, and secondly because it was felt that dissolving the cells directly in strong detergent minimized the possibility of degradation.

The DNAse treatment was usually for 15' but this was reduced to 1' in cases where the avoidance of nicking was especially important. A 1' exposure caused the DNA to break down to small pieces which precipitate in ethanol and sediment with tRNA.

Stage 3

In the first year of the work the B XIV zonal rotor was not available and so all the sucrose gradients were run on the SW 25.1 rotor of the Spinco model L centrifuge. The gradient was as supplied by M.Birnstiel:

Sodium acetate pH 5.0	20 mM
EDTA	1 mM
SDS	0.2%

5 - 40% sucrose.

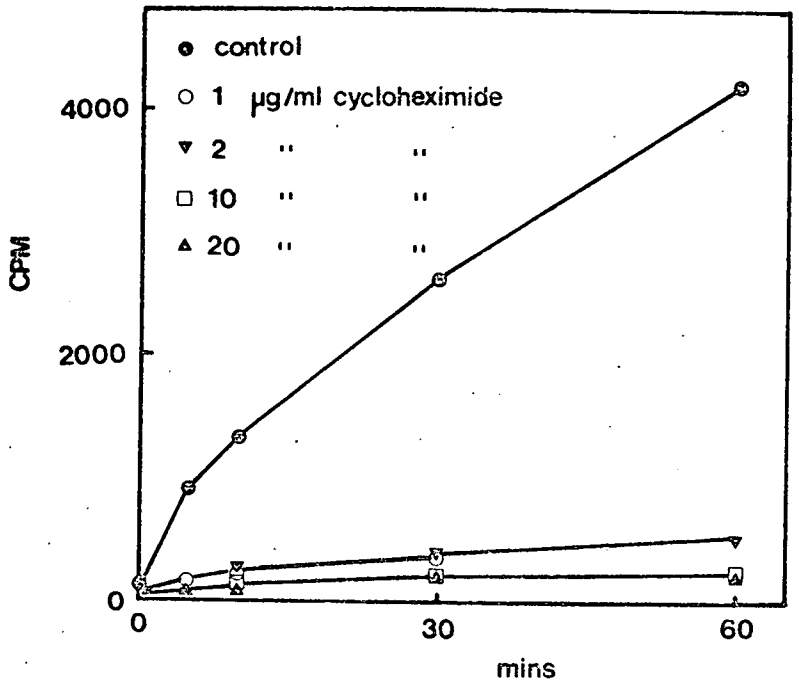


Fig.2.5 Effect of various concentrations of cycloheximide on the incorporation of ^{3}H-leucine by Xenopus kidney cells.

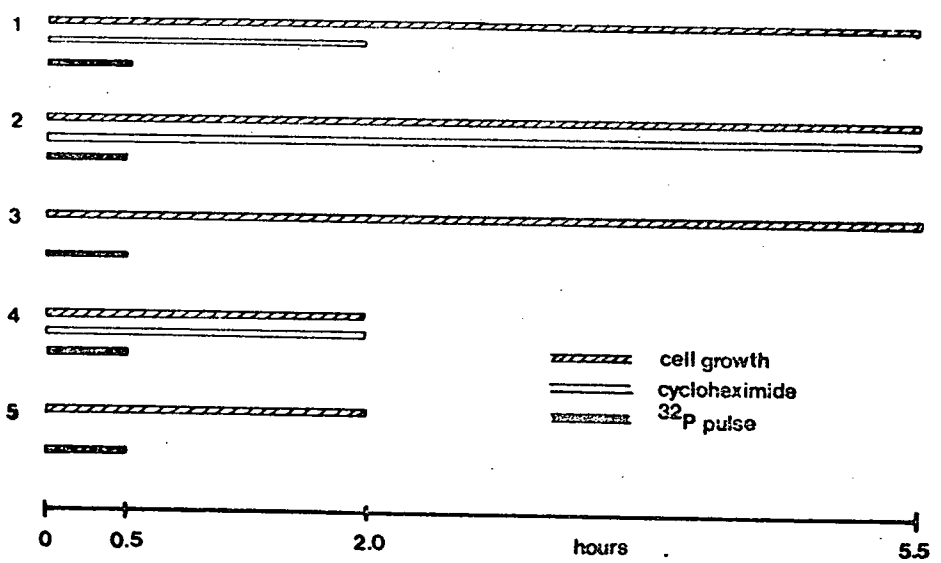


Fig.2.6 Regimes experienced by the five dishes of cells involved in the pulse chase experiment. See also Fig.2.7 overleaf.

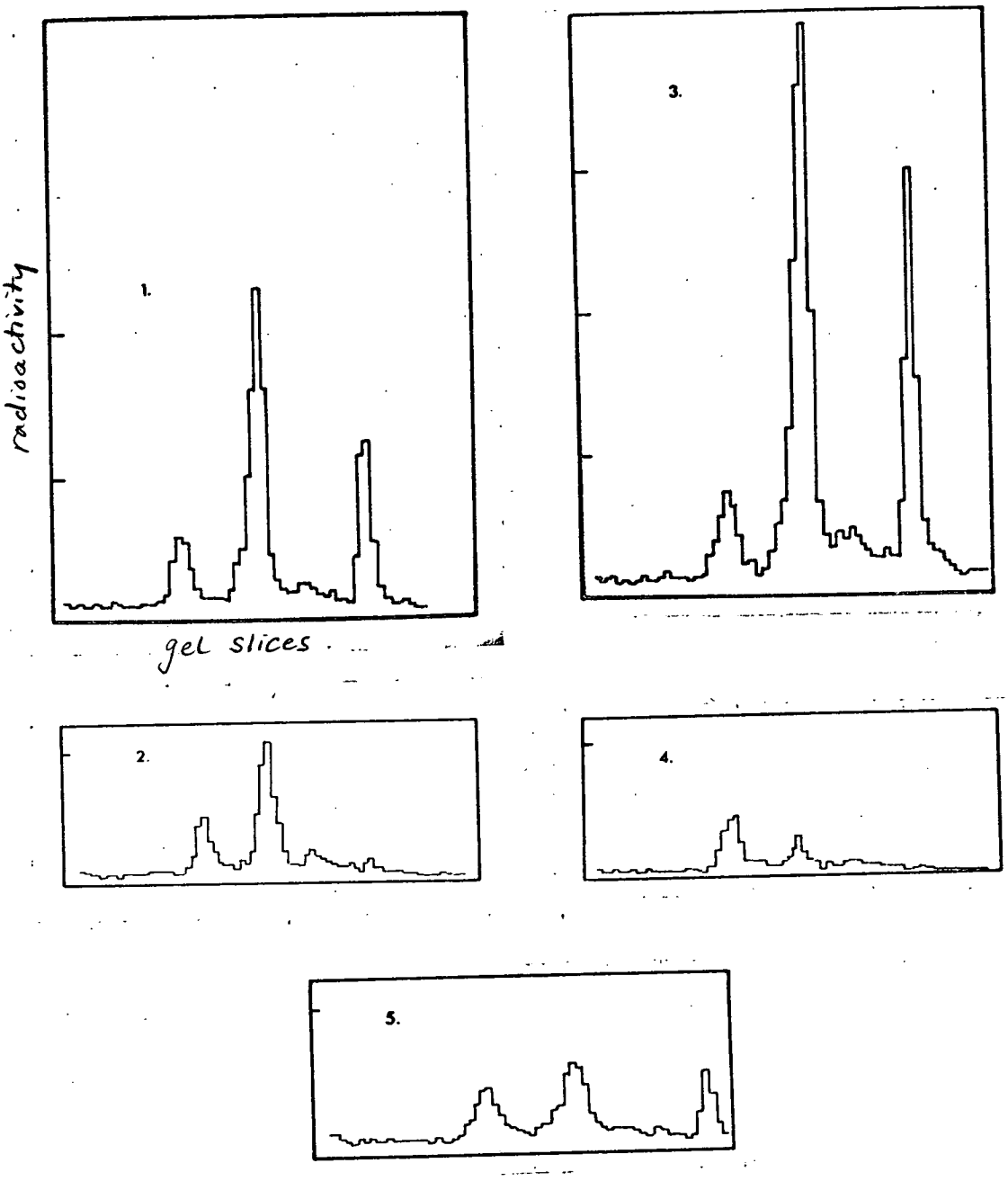


Fig. 2.7 Pulse-chase experiment. Dishes 5. and 3. were controls, dishes 4. and 2. were held in cycloheximide for 2 and 5.5 hours respectively, and dish 1. had the cycloheximide removed after 2 hours.

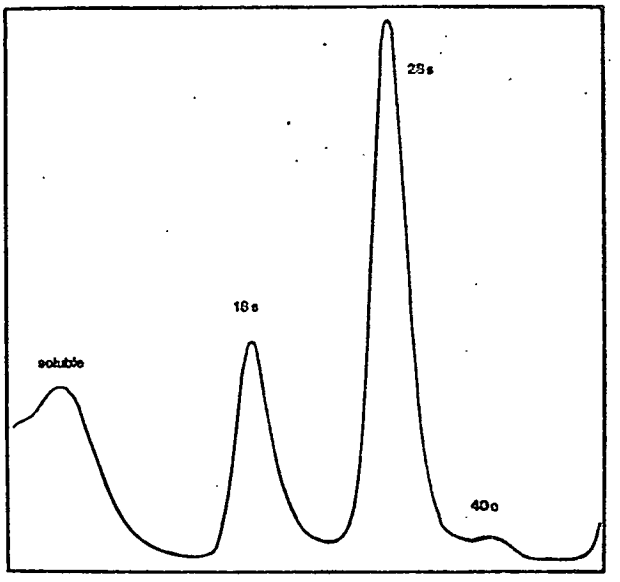


Fig. 2.8 Separation of Xenopus RNA species on a 5-40% sucrose gradient.

The gradients were run for 18 hours at 21.6K and 1 ml fractions were collected, the flow being monitored at 254 nm. A typical gradient is shown in Fig.2.8 above. The separation of 40s from 28s RNA is rather poor, but at this stage of the work purer preparations were not necessary.

When the zonal rotor became available, it was necessary to compute the best conditions for resolution of the RNA species. The relevant calculations are set out below and bear on the following questions:

i) What is a suitable approximation to an isokinetic gradient ?

ii) How is such a gradient to be prepared ?

In most sucrose gradients the zones move slower as they get further down the tube because the retardation caused by viscosity increases faster than the acceleration caused by the increased distance from the rotor

centre. However, a linear 5-20% gradient is usually approximately isokinetic and is for this reason favoured for molecular weight determinations (85). It was used for the 40s RNA preparation on the basis of the calculation set out in the table below. The approximately constant absolute velocity gave good prospects for complete separation from 28s RNA.

sucrose %	r cms	relative s-value	absolute velocity cm.sec.$^{-1}$x10^{6}
5	3.24	0.850	4.82
10	4.38	0.715	5.49
15	5.52	0.585	5.65
20	6.67	0.470	5.49

Outer radius of rotor = 6.67 cms.
Inner radius = 3.24 cms. (allowing for buffer overlay)

$$\text{Relative s value} = \frac{s}{s_{20,w}} = \frac{\eta_{20,w}\,(1-\bar{v}\rho)_{solv,T}}{\eta_{solv,T}\,(1-\bar{v}\rho)_{20,w}}$$

and since the runs were at 20°C, $\eta_{20,w} = \rho_{20,w} = 1$

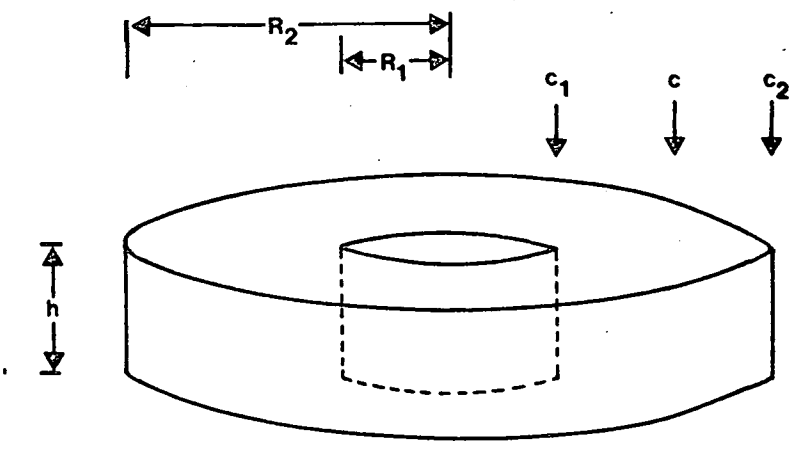
$\bar{v} = 0.53$ for ribosomal RNA (86)
η and ρ for sucrose solutions from (87).

Now, what form should such a gradient take if it is to be linear with radius inside the rotor ? If the dimensions of the rotor are as shown in Fig.2.9, and the effluent volume is represented by V, then:

$$r = R_1 \quad \text{and} \quad V = 0 \quad \text{at} \quad c = c_1$$
$$r = R_2 \quad \text{and} \quad V = V_f \quad \text{at} \quad c = c_2$$

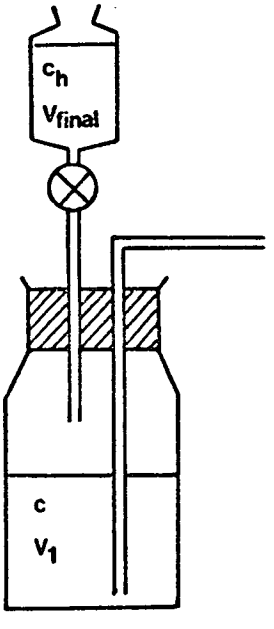
If the gradient is linear with radius, then:

/over



ZONAL ROTOR

Fig. 2.9



GRADIENT MAKER

Fig. 2.10

$$c = c_1 + \frac{r - R_1}{R_2 - R_1}(c_2 - c_1)$$

$$= r\frac{(c_2 - c_1)}{(R_2 - R_1)} + \left[c_1 - \frac{R_1(c_2 - c_1)}{R_2 - R_1}\right]$$

$$= r \cdot A + B$$

and for the B XIV rotor with a 5-20% gradient, A = 4.36 and B = -9.2.

The dependence of r on V can be deduced if the rotor is treated as a perfect annuloid with cylindrical volume elements:

$$dV = 2\pi r h dr$$

$$V = \pi r^2 h - R_1^2 h$$

$$r = \left[\frac{V}{\pi h} + R_1^2\right]$$

Substituting:

$$c = 4.36\left[\frac{V}{\pi h} + R_1^2\right]^{1/2} - 9.2 \quad \ldots(1)$$

It is rather difficult to design a gradient maker which produces a gradient of this particular mathematical form. But it was found that a very good approximation was produced by the apparatus described in ref.(88), Fig.2.10. For this gradient maker, the concentration of the effluent is given by: /over

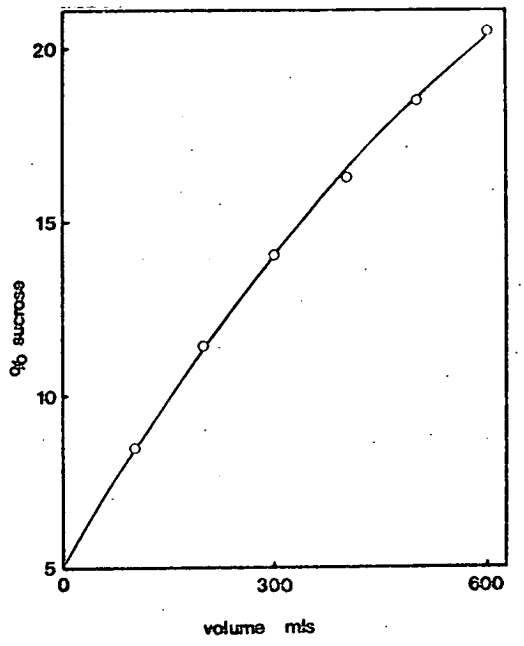


Fig 2.11 The curve represents the sucrose gradient required for linearity within the rotor, and the points represent the gradient produced by the apparatus in Fig. 2.10.

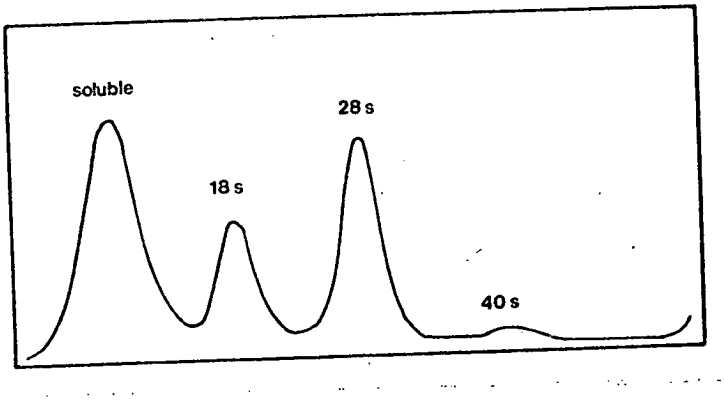


Fig 2.12 Separation of Xenopus RNA species by the zonal rotor.

$$c = c_h - (c_h - c_l)\, e^{-V/V_l}$$

and this gives good correspondance with equation (1) with the values c_h = 35% sucrose and V_l = 840 mls (Fig.2.11).

So all the gradients were made using these values and the samples were run as described above. It was calculated that a time of three hours at 45K would bring the 40s RNA two thirds of the way down the gradient, and the eventual time of centrifugation which was used was 3 hours 10 minutes. This method gave good resolution of soluble, 18s, 28s and 40s RNAs (Fig.2.12).

Stage 4

Both sucrose gradients and acrylamide gels separate mainly according to molecular weight, so they do not remove the heterogeneous RNA of the same molecular weight as the ribosomal precursor.

It has recently been shown that a certain proportion of HnRNA contains 3'poly A (89,90,91), and that poly A containing RNAs are retained on oligo-dT cellulose at high salt (92). For this reason, passage through an oligo-dT cellulose column was included as a step in the purification procedure.

The RNA which is retained on the column can be eluted with a low salt buffer (this is the same as the binding buffer without the NaCl). The 40s peak off the zonal gradient was used as the sample and a comparison was made between the retained and the non-retained RNA. Gel electrophoresis showed that the former was heterogeneous while the latter was mainly 40s (Fig.2.13).

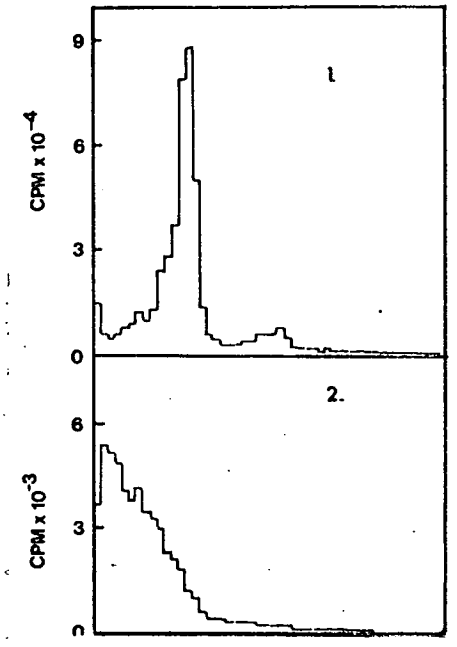


Fig 2.13 Oligo-dT cellulose fractionation of RNA from the 40s peak off the zonal rotor. 1) non-retained RNA 2) retained RNA.

An analysis of base composition showed that, as expected, the non-retained RNA is of high CG and the retained RNA of high A composition:

	A	U	C	G	
Non-retained	22	17	31	30	%
Retained	28	22	26	24	%

Stages 5 & 6

At this stage, a 40s RNA preparation still usually contained a certain amount of 28s RNA which probably sedimented as an aggregate in the zonal rotor. An attempt was made to separate the two using a 5-20 % gradient in the Spinco SW.50 rotor. This rotor goes up to 50K and so a run can be completed in 90 minutes and diffusion be thereby minimised. However, it seemed that the resolving power was not good enough (Fig.2.14).

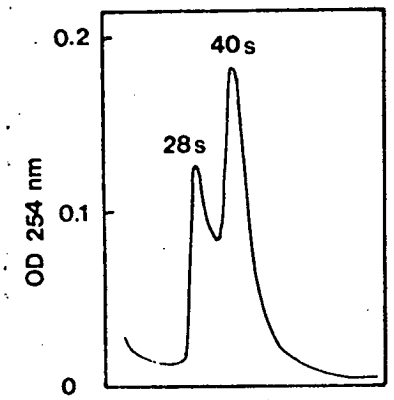


Fig 2.14 Failure to separate 40s from 28s RNA by sucrose gradient sedimentation.

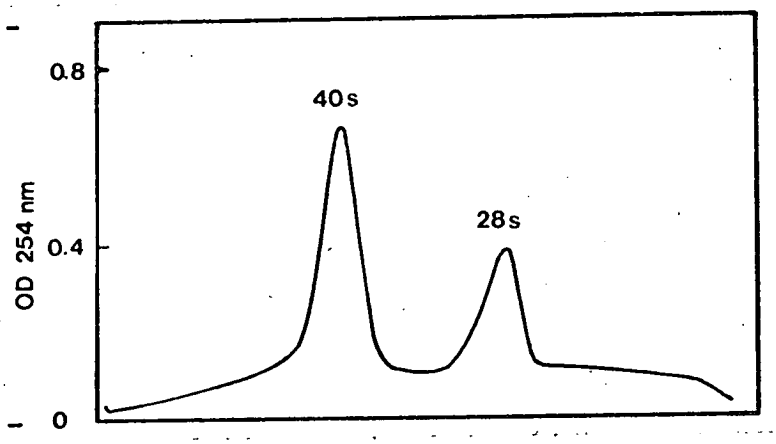


Fig 2.15 Separation of 40s from 28s RNA by polyacrylamide gel electrophoresis.

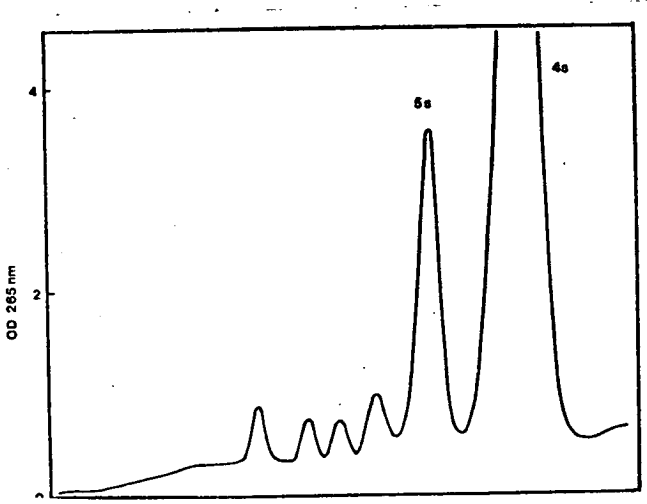


Fig 2.16 Separation of 5s from 4s RNA by polyacrylamide gel electrophoresis.

So, polyacrylamide gel electrophoresis was used routinely. This gives good resolution but has the disadvantage that a certain amount of oligo-acrylamide comes out of the gel along with the RNA.

40s RNA was prepared with 2.4% gels (Fig.2.15), and 5s RNA was similarly purified from the first zonal rotor peak using 7.5% gels (Fig.2.16). It was thought possible that some of the four minor peaks on the 7.5% gels might be messenger RNAs, but none of them was found to stick to oligo-dT cellulose under the conditions described above.

Recovery of the RNA from the gels was for some time carried out by the method in use in the laboratory. This involved shaking the slices for some hours in 0.6 M lithium acetate, 0.5% SDS and then pelleting the RNA from solution with an overnight centrifugation at 40-50K. This procedure takes two days if the elution is done overnight and does not give more than 50% yield of 40s RNA. It has the advantage that single gel slices can be eluted.

The method described in the protocol, in which the RNA band is electrophoresed into a dialysis bag, was found to be much quicker and to give an almost quantitative recovery. Variants of this technique which involved the use of sucrose cushions in the bag were found to have nothing in their favour.

Discussion

The inhibition of processing by cycloheximide presumably indicates that newly synthesized protein is required, whether or not these are ribosomal proteins

is unknown. Accumulation of the precursor RNA does not occur in HeLa cells and reaches a maximum after two hours in Xenopus. The simplest explanation of this is a feedback control of transcription by the amount of precursor. But there must be other controls on rRNA synthesis as witnessed for example by the finding that liver contains much less pre-rRNA than tissue culture cells, a difference which is presumably related to the rate of cell division and the requirement for new ribosomes. Such controls of rRNA synthesis are discussed by Perry (93).

Chapter 3

THE "HETEROGENEITY" EFFECT IN 40s RNA

Introduction

The stimulus to this part of the work was the heterogeneity of 40s RNA revealed by gel electrophoresis (78). If the RNA from various slices across the peak was eluted and rerun, then it ran in the position from which it was taken. This shows that the molecules from different parts of the peak differ in some way and cannot rapidly interconvert. There seemed to be four possible explanations:

i) 40s RNA is polydisperse in the same sense as a synthetic polymer, and contains a mixture of components of different lengths. Either the initial transcripts are different or this could be a result of early processing reactions.
ii) Although all of the molecules are the same length, they differ slightly in sequence because the ribosomal genes are not all identical.
iii) The differences are of secondary structure and represent a series of conformations adopted by the molecule in the various nucleolar RNP particles.
iv) The differences are of secondary structure but are of no functional significance.

The experiments described in this chapter depended on the use of formamide, which has been shown to denature RNA (94). There were two main types of experiment. In both, ^{32}P labelled RNA eluted from single gel slices was rerun in the presence of ^{3}H labelled marker RNA. In the first type of experiment, both RNAs were denatured just before the rerun, and in the

second type the ^{32}P labelled RNA was denatured before the first run.

Formamide Denaturation and Gel Electrophoresis

These experiments involved the use of RNA pulse- or long labelled with ^{32}P and ^{3}H-uridine. The techniques of cell growth, labelling, RNA preparation, gel manufacture and elution are all described above. Sterile precautions were extremely important for these preparations, which were all of "total RNA" and performed on a small scale.

The formamide was deionized by passage through mixed Amberlite IR-120/IRA-400. The conditions used for the denaturation of RNA were 70% formamide at room temperature for a few minutes, the RNA being dissolved in "E" buffer.

Denaturation and Heterogeneity

When the helical regions of RNA are unwound this causes changes in various physical properties. It also causes breaks in the chain to be exposed, because molecules containing such breaks ("nicks") will fall apart.

The denaturation of RNA as measured by hyperchromicity was determined for a range of formamide concentrations at room temperature using 28s RNA (Fig.3.1), and a formamide concentration of 70% was judged to be sufficient for complete denaturation. Whether it really is sufficient or not will be discussed below.

The release of 5.8s RNA from the 28s was studied

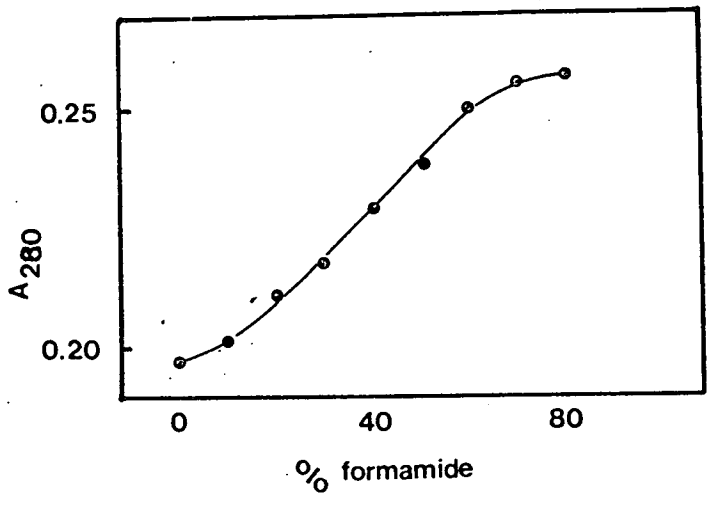


Fig.3.1. Denaturation of 28s RNA by formamide. This hyperchromicity curve was determined by adding RNA samples to different concentrations of formamide. The optical density was measured at 280 nm rather than 260 nm and the formamide absorption was subtracted.

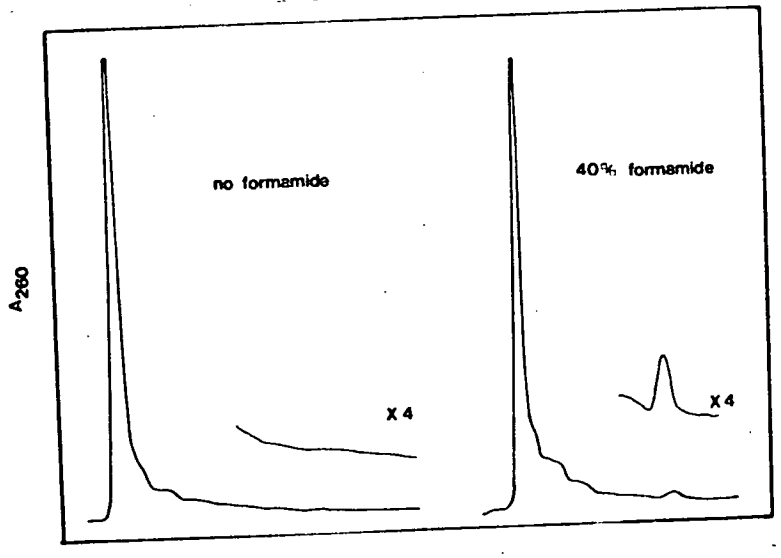


Fig.3.2. Denaturation of 28s RNA by formamide. The release of 5.8s RNA is revealed by polyacrylamide gel electrophoresis.

by running the samples on 2.4% gels for one hour (Fig.3.2). It was found that this fragment was released at a formamide concentration of 20% or more. In other words, although complete denaturation required 70% formamide, various "partial reactions" of denaturation could be completed at much lower concentrations.

Denaturation of subfractions of 40s RNA

RNA eluted from slices taken from either side of the 40s peak reran in the same position relative to the ^{3}H marker RNA. But when this RNA, together with the marker, was denatured in formamide, the labels were found to run coincidently (Fig.3.3).

So, that which causes the effect can be abolished by denaturation. This rules out number ii. from the possibilities listed above. Number i. could still be true if all the molecules contained a nick such that they were "tailored" to the same size by denaturation, but the third and fourth possibilities seem more likely.

A decision between these two was made by means of the second type of experiment. If the ^{32}P-RNA is denatured before the first run, then the molecules should become homogeneous if they have been tailored or if they were in special conformations brought about by the effect of other molecules in vivo. But if the fourth possibility is correct and the heterogeneity is caused by spontaneously forming but non-interconvertable conformations, then it should remain because the molecules will coil up again when they have migrated out of the formamide.

This was found to be the case (Fig.3.4), and so it was concluded that the fourth explanation was the

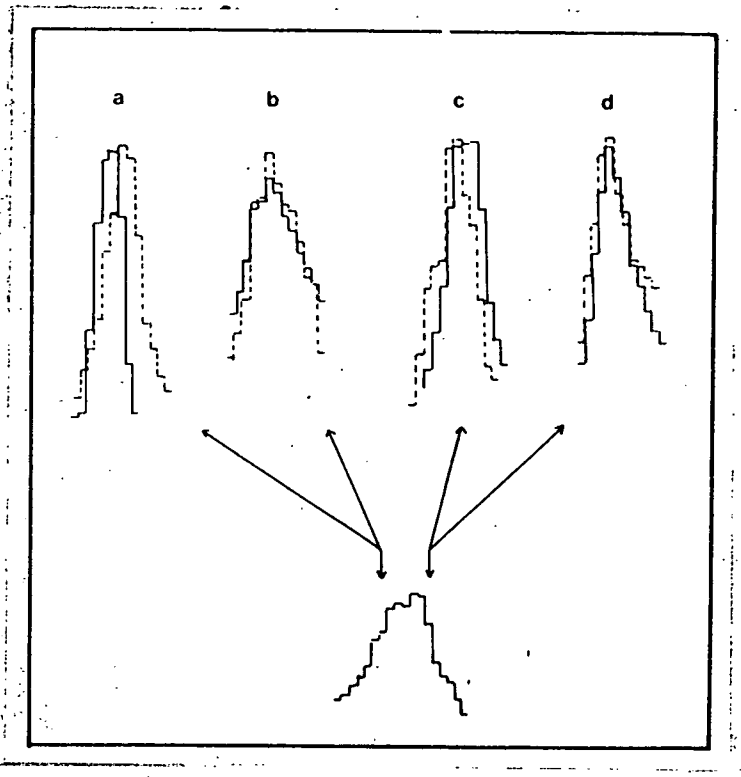


Fig.3.3. Abolition of the heterogeneity of 40s RNA by denaturation. The first gel peak is shown below and RNA from the indicated slices was rerun. In the four upper gels the dashed histograms represent the ^{3}H-marker RNA. The RNA in gels a. and c. was unmelted and that in gels b. and d. was melted.

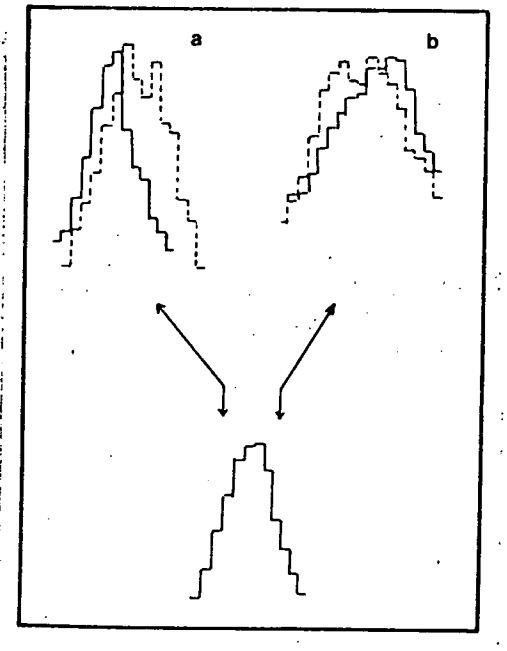


Fig.3.4. In this experiment the RNA was denatured before the first run. It can be seen that the heterogeneity remains.

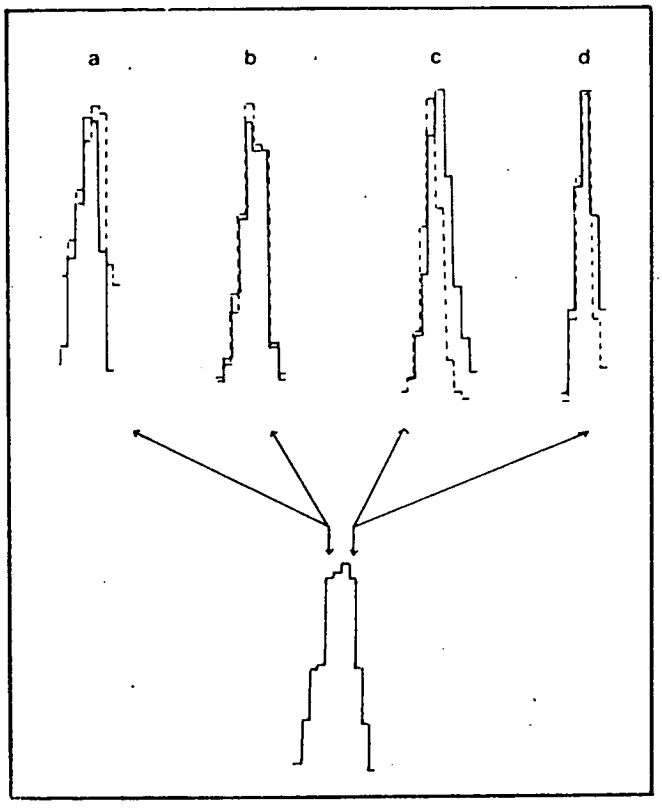


Fig.3.5. The heterogeneity effect in 28s RNA. This RNA can be seen to show a slight but definite heterogeneity (a. & c.) which is also abolished by formamide (b. & d.).

correct one. This means that the effect should be shown by all RNAs, although one might expect the number of stable and non-interconvertable forms to increase with molecular weight. Accordingly, the first experiment was repeated with 28s RNA and it gave the same result: the RNA showed slight heterogeneous behaviour which was abolished by denaturation (Fig.3.5).

Discussion

The secondary structure of RNA in neutral solution is thought to involve bonding of anti-parallel strands

with A-U and C-G base pairs (95). It seems intuitively possible that when the molecules are synthesized, or when they are transferred out of a denaturing environment, the base paired loops do not form identically in all molecules. Once formed, they would be metastable if interconversion involved unfolded intermediates of a significantly higher free energy. The existence of such an array of metastable states would be a plausible explanation of the results described above, because gel mobility is known to depend to some extent on secondary structure (96). It can be predicted that there would be more metastable states the longer the molecule, in which case the feature of 40s RNA which makes it appear heterogeneous is simply its length.

The existence of metastable conformations in RNA is already known from studies of titration hysteresis between pH 3 and 7 (97,98). These,however, are not thought to involve the normal base pairs but parallel runs of A-AH^+ or C-CH^+ (98) or a variety of G-G base pairs (99), so the non-occurrence of titration hysteresis in the pH 7-8 range does not necessarily mean that there are no metastable states under these conditions. This conclusion about RNA structure is rather different from the assumption of Fresco et.al. (100) who wrote in 1960: "RNA takes up its most stable conformation as a result of a lateral mobility which permits exploration of most competitive conformations."

Since these experiments were carried out, rumours have been circulating to the effect that the conditions used were not sufficient to bring about complete denaturation. This may be so, but it seems that the conditions were stringent enough to abolish the heterogeneity. Presumably the last loops to unwind are the longest and most CG rich, and it might be that these

form very fast in renaturing conditions and are thus the same for all molecules.

Chapter 4

INVESTIGATION OF 5' ENDS OF 18, 28, AND 40s RNAs AND OF OTHER ALKALINE HYDROLYSIS PRODUCTS

Introduction

As was explained in the first chapter, an examination of the 5' termini of the Xenopus rRNAs was carried out for three reasons: to find out whether or not the 40s RNA was polydisperse, to find if the 40 and 28s RNAs had the same 5' end or methylated sequence, and to find whether the 40s RNA bore a 5' triphosphate characteristic of a primary transcription product.

It was felt that the best way to approach these questions was the rather old technique of alkaline hydrolysis coupled with chromatography in 7 M urea on DEAE Sephadex (101, 102, 103). Alkaline hydrolysis will break down an RNA sequence as follows (104, 105, 106):

$$pNpN.....pNmpN.....pN_{OH}$$

$$\downarrow OH^-$$

$$pNp + Np + NmpNp + N_{OH}$$

$$-4 \quad -2 \quad -3 \quad 0$$

....expected negative charge at neutral pH.

By far the majority of counts in a ^{32}P labelled sample appear in the 5' monophosphates. A small percentage appears in dinucleotides which are methylated at the

2'-O-position of the 5' residue. This methylation prevents the formation of the cyclic phosphate intermediate of hydrolysis. The 5' terminus appears as a nucleoside-3',5'-diphosphate, and the 3' terminus disappears in a ^{32}P labelled sample because it forms an unlabelled nucleoside. If the 5' end bears a triphosphate, the hydrolysate will contain a product of the type pppNp which has a negative charge of about -6 at neutral pH.

DEAE columns with 7 M urea have been widely used for separating oligonucleotides. The urea suppresses the effects of non-ionic interaction between the nucleotides and the column medium. So a linear salt gradient should elute species in order of their negative charges. Commonly, a digest of yeast RNA with pancreatic RNAse is used as an optical standard, because it provides a regular pattern of isostich peaks, each with one more negative charge than the previous one.

For the unambiguous detection of 5' termini, quite a large amount of radioactivity is required in the sample. For example, consider the 40s molecule which contains about 8200 phosphate groups. With a sample of one million cpm we shall obtain 122 cpm per phosphate group. In practice, about 10^6 cpm were loaded onto the column and the relevant fractions were counted for 10'. So a tetraphosphate should give 4880 counts and be easily detected.

Peaks which were likely candidates for 5' terminus derivitives were subjected to two tests:

i) Chromatography with pNp standards.

ii) Treatment with alkaline phosphatase followed by chromatography. Phosphatase causes end group derivitives to lose all their label as phosphate,

while oligonucleotides lose only some of their label:

$$pNp \longrightarrow N + 2P$$

$$pppNp \longrightarrow N + 4P$$

$$NpNpNp \longrightarrow NpNpN + P$$

Column and Thin Layer Chromatography

RNA Preparation

This was described in Chapter 2.

Alkaline Hydrolysis

The RNA sample, containing at least 10^6 Cpm, was incubated in about 1 ml of 0.2 M KOH at 37° for 18 hours. Then 5% perchloric acid was added dropwise until the solution was neutralized, as determined by spotting small quantities onto pH paper. The solution was left in ice for 15' to allow the $KClO_4$ to precipitate. The precipitate was removed by centrifugation and the desalted sample was mixed with the optical marker and diluted with 8 M urea to a final concentration of 7 M.

Preparation of Optical Marker

RNA from Torula (Sigma, Grade VI) was used. It was found to give a rather unsatisfactory isostich pattern unless it was further purified. So it was subjected to phenol extraction and repeated precipitation with ethanol. The marker was made by incubating the RNA with 1/20th of its weight of pancreatic RNAse

for 1 hour at 37° in 10 mM Tris Cl pH 7.4, 1 mM EDTA.

DEAE Columns

Initially, Whatman DE-52 cellulose was used in an 0.9 x 30 cm column. This was not very successful and eventually "microcolumns" of 0.3 x 20 cms containing DE-Sephadex A 25 were found to be most useful. The DE-Sephadex was precycled in 0.1 M HCl and NaOH and equilibrated in the starting buffer. The buffer was initially 20 mM Tris Cl pH 7.8, 1 mM EDTA, 7 M urea, but later the EDTA was omitted for reasons which will be described shortly. The urea (BDH Analar) was made up to 8 M in distilled water and this solution was purified of charged impurities by serial passage through a column of DEAE cellulose and phosphocellulose.

The microcolumns took about six hours to run. The gradient was 0-0.4 M NaCl and 100 one ml fractions were collected. This system separated the first six isostichs quite reproducibly. The effluent optical density at 254 nm was recorded continuously with an ISCO monitor. The fractions were collected inside small plastic tubes which fitted inside scintillation vials. They were counted by Cerenkov radiation (55% gain, full window) on a Packard Model 3320 scintillation counter.

Removal of Salt and Urea

It was necessary to remove the salt and urea from the column fractions which contained material to be subjected to further analysis. Three methods were tried in order of recommendation by the resident authorities, and only the third was any good. The first was a batch procedure which involved the absorption of the sample onto a small amount of DE-52, and it

was rather unreproducible. Either the radioactivity could not be completely absorbed, or it could not be eluted. In the second method, the counts were absorbed onto activated charcoal and eluted with ethanol, and this did not work because the counts always came off in distilled water. The successful method was a modification of that described by Rushizsky and Sober (107). The solution was diluted five fold with distilled water and run slowly through a 1 ml column of DE-52. It was washed with 30 mls of 20 mM ammonium carbonate pH 8.0, and eluted with 30% triethylammonium carbonate (TEC) pH 9.2. Five drop fractions were collected and counted by Cerenkov radiation. Those fractions which contained the radioactivity were spotted onto a glass backed polythene sheet and dried down in a vacuum dessicator. The residue was redissolved in distilled water and dried down again a few times.

The White Sludge

Although the Rushizsky-Sober technique gave quantitative recovery of the radioactivity, it did not at first seem to remove all the salt. There was always a considerable residue left after removal of the TEC. Oddly enough, this was not very soluble in water and quite a lot of it could be removed by mixing the residue with a little water and centrifuging. But enough remained to stop the samples running properly on the thin layers. In addition it became clear after a while that it was inhibiting the alkaline phosphatase which was being used to test for terminal phosphates.

What could the sludge be ? How could it be got rid of ? Whatever precautionary measures were taken, the sludge always seemed to be present in large amounts. Eventually one suspicion was confirmed when it was found that the addition of zinc sulphate to a mixture of

alkaline phosphatase, p-nitrophenyl phosphate and sludge caused it to turn yellow. The material was EDTA from the column buffer. It was concentrated by the desalting procedure because of its multiple charge, and it inhibited the phosphatase by chelation of zinc. The amounts encountered could be explained because five mls of pooled fractions at 1 mM EDTA would contain over two milligrams of the material.

pH 2.7 Columns

For some purposes, DEAE Sephadex columns were run at acid pH. Here the buffer consisted of 7 M urea titrated to pH 2.7 with HCl, and the gradient was 0-80 mM sodium chloride. The pH used is in between the first and second pKs of phosphate, so each phosphate group bears one negative charge whether its situation is internal or terminal. Also the nucleoside bases bear partial positive charges, so there is separation by base composition:

Phosphate	Charges at pH 2.7	
pK_1 1.0-1.6	C	+1.00
pK_2 5.8-6.6	A	+0.95
	G	+0.30
	U	0 (ref.108)

Enzyme Treatments

The reactions were carried out in liquid drops on polythene sheets, evaporation being prevented by a plastic vial cap stuck down with vacuum grease.

For phosphatasing, the samples were dissolved in 5 µl of 20 mM TEC pH 8, 5 mM $MgCl_2$ and treated with

0.5 μg of E.coli alkaline phosphatase (Whatman) at 37° for 30'.

For complete venom phosphodiesterase digestion, the buffer was 5 μl of 50 mM Tris acetate pH 8.2, 5 mM $MgCl_2$. Incubation was for 3 hours at 37° with 1 μg of enzyme (Worthington).

Thin Layer Chromatography

This was carried out on commercial thin layers of cellulose impregnated with polyethylene imine (PEI-cellulose, Camlab) (109,110). These were prepared for use by i) wetting in distilled water, ii) washing in 10% NaCl, iii) washing in water, iv) washing in 2 M pyridine formate pH 2.2, v) rinsing in water and allowing to dry. Chromatography was performed in the upwards direction on 20 cm strips, sometimes a paper wick was attached to the top of the strip to draw the solvent further. The solvent was 2 M sodium formate or 1 M pyridine formate pH 3.5. All thin layer separations in this and subsequent chapters were accompanied by a mixture of dyes consisting of xylene cymol FF, Orange G and Acid fuschin. The chromatograms were visualized by autoradiography with X ray film. The development of this technique and the synthesis of the standards is described in the next chapter.

Results

Total RNA

The first question was whether triphosphate ends could be detected in total pulse-labelled RNA, as is the case in E.coli (111). The cells were labelled for

one hour and the total RNA prepared as described above. It was dialysed for 48 hours against several changes of 0.15 M sodium acetate, 0.5% SDS to remove low molecular weight contaminants.

The column fractionation of the alkaline hydrolysate is shown in Fig.4.1. The peak labelled "A" was suspected to contain the 5' derived tetraphosphates because of its position (112). The material was desalted and divided into two portions. One was phosphatased and both were chromatographed on PEI-cellulose. As shown in Fig.4.2, the material is immobile in the solvent used and all its phosphate is labile to the enzyme, both of which properties are consistent with it being tetraphosphate.

It was concluded that at least some RNA species in Xenopus bear 5' triphosphates and that they were detectable by means of the techniques used. The result is not surprising because 5' triphosphates have, for example, been shown to exist in rat liver HnRNA (113).

5s RNA

5s RNA from all eukaryotes is reported to be a mixture of molecules containing one, two and three 5' terminal phosphates (114). Xenopus 5s RNA seemed therefore to be an ideal control for the technique. The 5s RNA was prepared as described above, hydrolysed and run on the column (Fig.4.3). Three small peaks were found to elute above the mononucleotides. The results from the phosphatasing and chromatography were:

Peak	Chromatography	Phosphatase
I	streaked	loses all
II	with pGp	"
III	immobile	" (Fig.4.4)

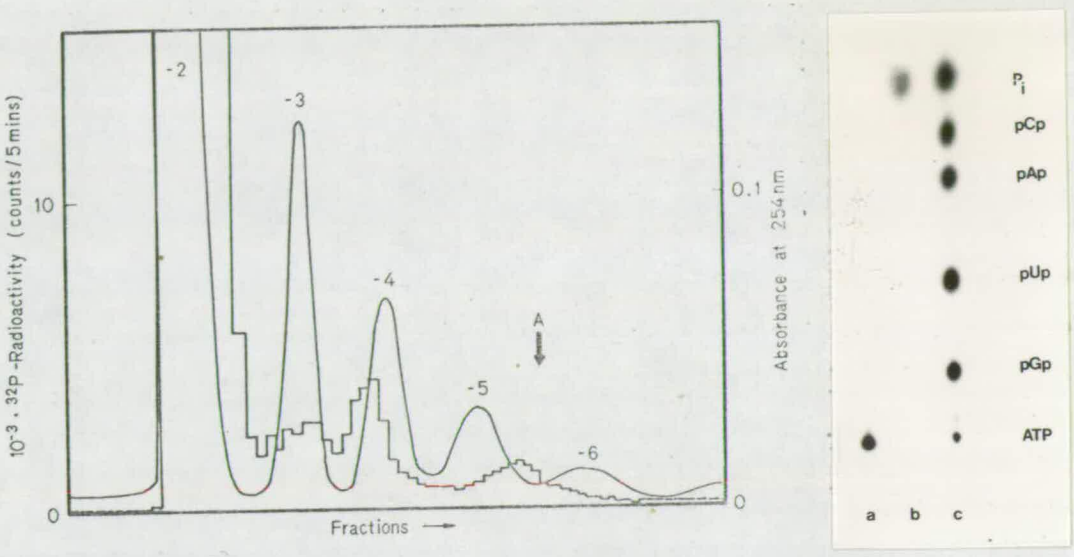


Fig.4.1. DE-Sephadex column of total RNA hydrolysate. The continuous line shows the optical density trace of the isostich markers, and the histogram shows ^{32}P radioactivity.

Fig.4.2. PEI-cellulose chromatography. a) peak "A", b) peak "A" after phosphatase, c) standards.

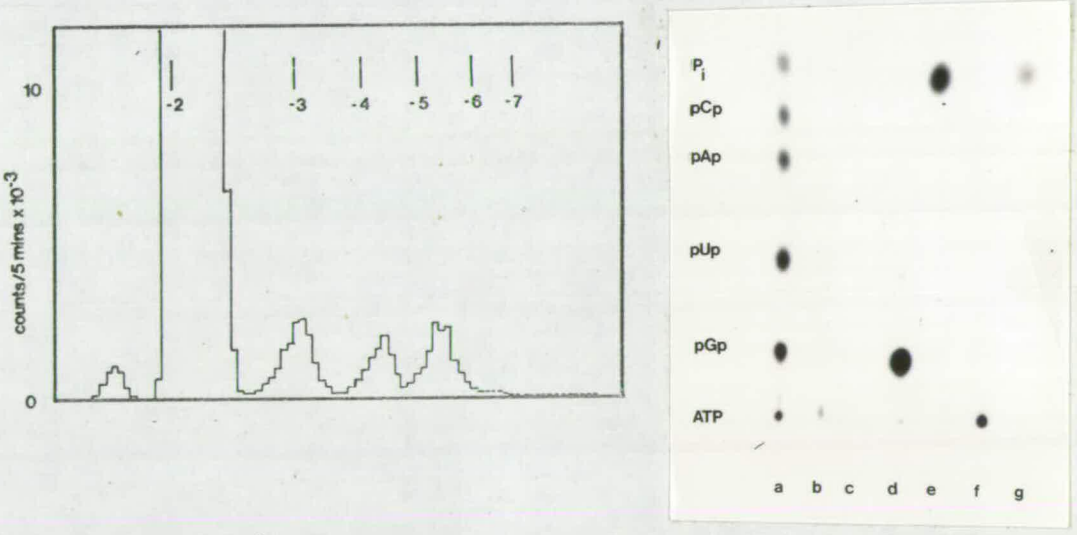


Fig.4.3. DE-Sephadex column of 5s RNA hydrolysate. The bars show the positions of elution of each isostich, with the negative charge, and the histogram shows ^{32}P radioactivity.

Fig.4.4. PEI-cellulose chromatography. a) standards, b) & c) peak I before and after phosphatase, d) & e) peak II before and after phosphatase, f) & g) peak III before and after phosphatase.

It is possible that the three peaks are pGp, ppGp and pppGp, but the tests rather suggest that the second is pGp and the third ppGp. The technique does not distinguish a ppNp from a pppNp, but this is not very serious since both indicate that the molecule is a primary transcription product. It should be noted that the end group determination is correct, since the 5' residue was already known to be G (31).

18s RNA

The column separation of the 18s hydrolysate is shown in Fig.4.5. There was a sizeable peak eluting with the second isostich (-3) and a small one eluting with the third (-4). Since ribosomal RNA was known to contain 2'-O-methylations, the -3 peak was thought to consist of methylated dinucleotides of the type NmpNp. This was confirmed by a methyl labelling experiment. RNA labelled with ^{32}P was mixed with RNA labelled with (^{3}H-methyl)-methionine before the zonal rotor separation. The 18s RNA was hydrolysed and chromatographed as before and the column fractions counted for both isotopes (Fig.4.6.). There is some ^{3}H label in the mononucleotides, but it is not certain that this is significant because of the great excess of ^{32}P in this peak. The -3 peak was strongly labelled with ^{3}H and the -4 peak unlabelled. So this strongly suggests that the -3 peak consists of methylated dinucleotides and that the -4 peak was a nucleoside-3',5'-diphosphate from the 5' end. The proportion of radioactivity in the -4 peak, averaged over several experiments, indicated 2.3 moles of phosphate which is about right for one mole of pNp.

The material from the -4 peak was desalted and run on a DE-Sephadex column at pH 2.7. It ran as a single peak, apart from a few counts in the runoff,

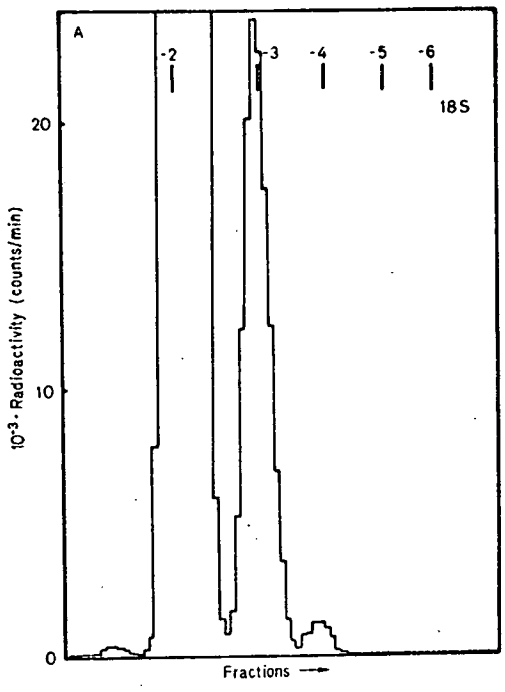


Fig.4.5. DE-Sephadex chromatography of 18s RNA hydrolysate. The bars show the position of the isostichs with their net charges, and the histogram shows ^{32}P radioactivity.

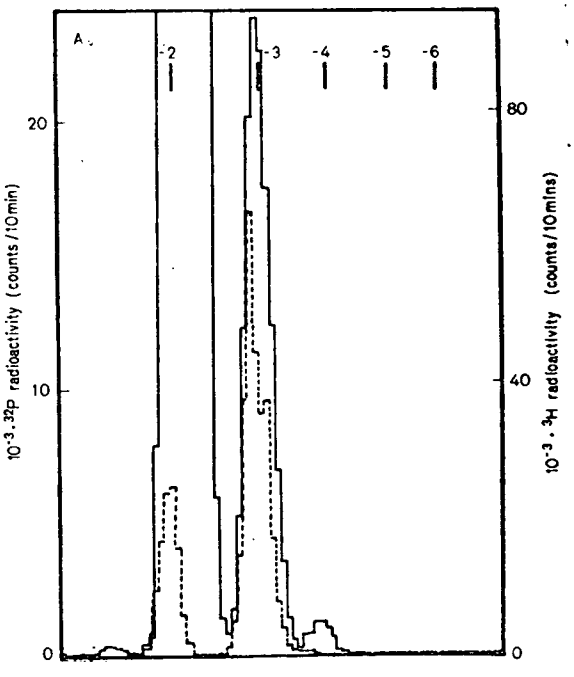
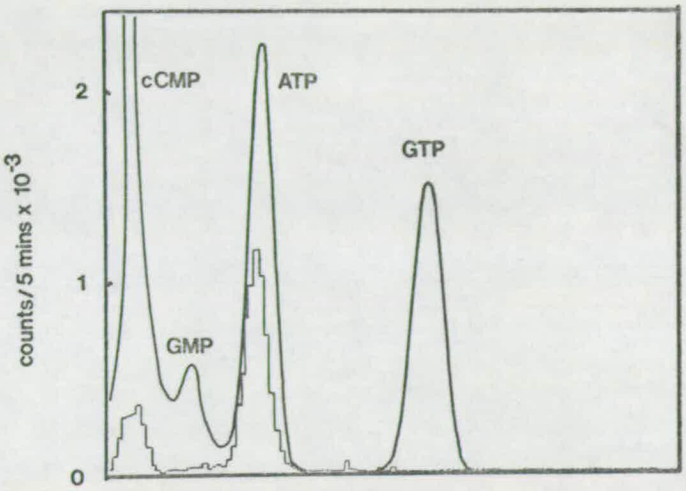


Fig.4.6. The same with the ^{3}H radioactivity from a (^{3}H-methyl)-methionine label. The -3 peak is labelled with ^{3}H while the -4 peak is not.



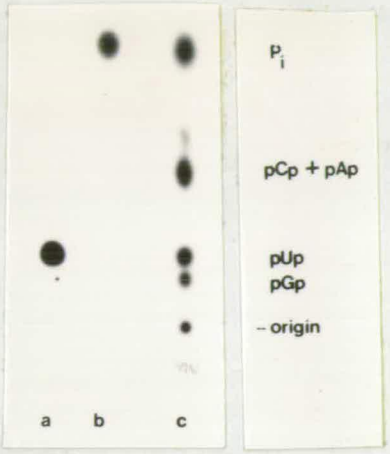


Fig.4.7. DE-Sephadex chromatography at pH 2.7 of the -4 peak from the 18s hydrolysate. The bulk of the counts coelute with ATP at this pH.
Fig.4.8. PEI-cellulose chromatography. a) -4 peak from 18s RNA, b) the same after phosphatase, c) standards.

and eluted along with ATP (Fig.4.7). At pH 2.7, ATP has a charge of -2, and the only nucleoside-3',5'-diphosphate which would have this charge is pUp (see p56). An analysis of the material showed that it ran coincidently with pUp on PEI-cellulose and that all its phosphate was labile to phosphatase (Fig4.8). So it was concluded that the 5' terminus of the 18s RNA was pU.

28s RNA

The column separation of the 28s hydrolysate is shown in Fig.4.9. Here, there was again a relatively large -3 peak, and this time two small peaks of charge -4 and -5. The number of phosphate groups per mole was estimated at 3.4 for the -5 peak and 7.8 for the -4 peak. All of these three peaks were found to be labelled by (^{3}H-methyl)-methionine (Fig4.10). There was no peak in a position appropriate to a pppNp.

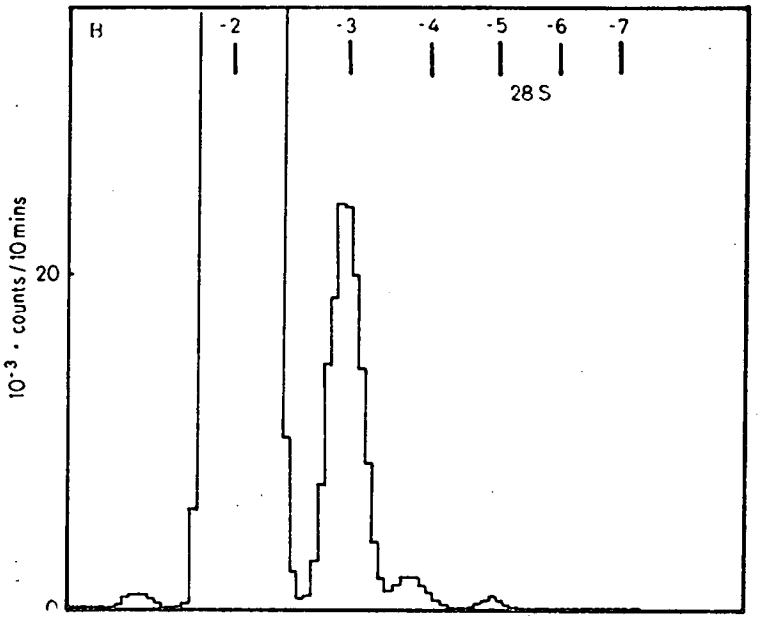


Fig.4.9. DE-Sephadex column of 28s RNA hydrolysate. The bars show the positions of the isostichs and the histogram the ^{32}P radioactivity.

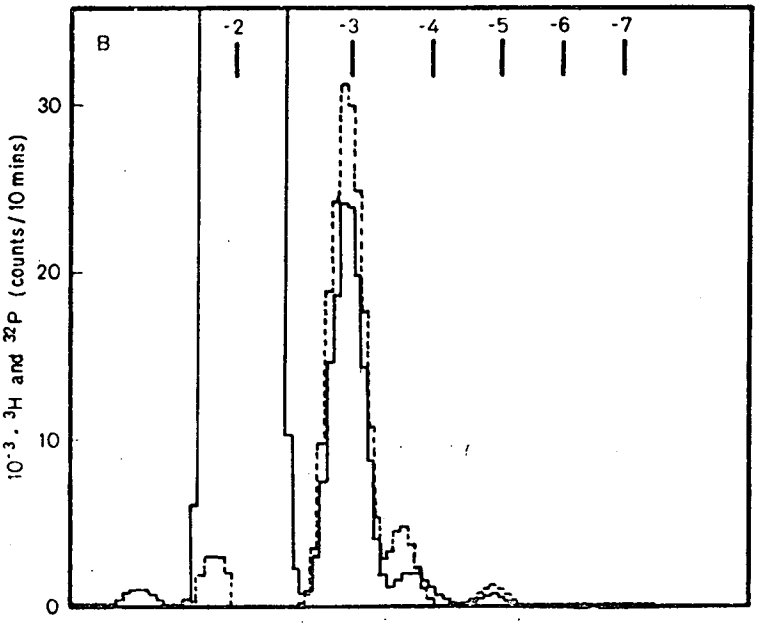


Fig.4.10. The same, with methyl-^{3}H counts indicated by the dashed histogram.

The -5 peak was of some interest because it could have been one of several things.

		Me/P	number of phosphates
1.	ppNp	0	3
2.	pNmpNp	0.33	3
3.	NmpNmpNmpNp	0.75	4

The first would clearly be a 5' end derivitive , the second would be a 5' end derivitive of the type described by Choi and Busch (81) for rat 28 and 40s RNAs, and the third would be a methylated stretch of unprecedented length from a naturally occurring RNA.

The material was shown to be a single component by electrophoresis on DEAE paper at pH 1.9. The estimated number of phosphate groups, 3.4, was consistent with any of the possibilities. The methyl labelling, however, excluded the first. The ratio of Me/P, taking the alkali stable dinucleotides as 0.5, was 0.63. This was closer to the third than to the second possibility but not really conclusive.

The crucial experiment consisted in rerunning the material on the DE-Sephadex column after digestion with alkaline phosphatase. This is the "shift out" technique of Tomlinson and Tener (115) which is especially useful for distinguishing between terminal and internal fragments:

ppNp $\longrightarrow$ N + 3P
-5 0 -2 ...net charge at pH 7.8

pNmpNp $\longrightarrow$ NmpN + 2P
-5 -1 -2

/over

$$\underset{-5}{\text{NmpNmpNmpNp}} \longrightarrow \underset{-3}{\text{NmpNmpNmpN}} + \underset{-2}{\text{P}}$$

In fact peaks were obtained in the -2 and -3 positions (Fig.4.11) and this is very good evidence that the material is a tetranucleotide and not a methylated 5' end. It was thought possible, particularly before the methyl labelling experiment was done, that the tetranucleotide might result simply from incomplete hydrolysis. But when it was desalted, rehydrolysed for 18 hours and rerun on the column, all the radioactivity was still at the -5 position (Fig.4.12). This should also exclude the structure NmpNpNmpNp which might have a relatively resistant internal phosphodiester bond.

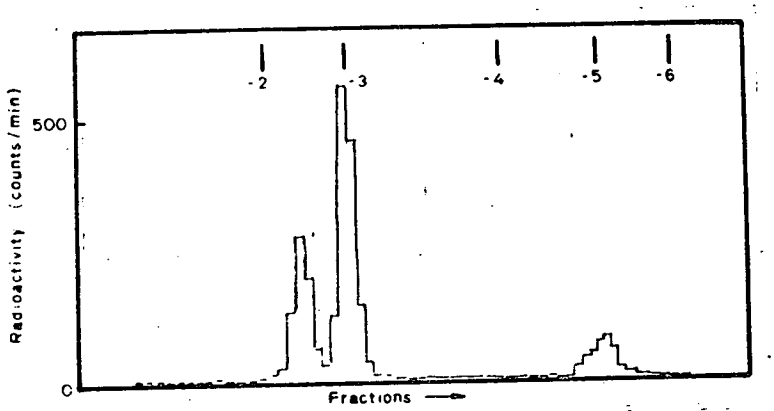


Fig.4.11. "Shift out" experiment. Treatment of the -5 material with phosphatase gives peaks at the -2 and -3 positions showing that the material was a tetranucleotide. In this experiment the phosphatasing was incomplete.

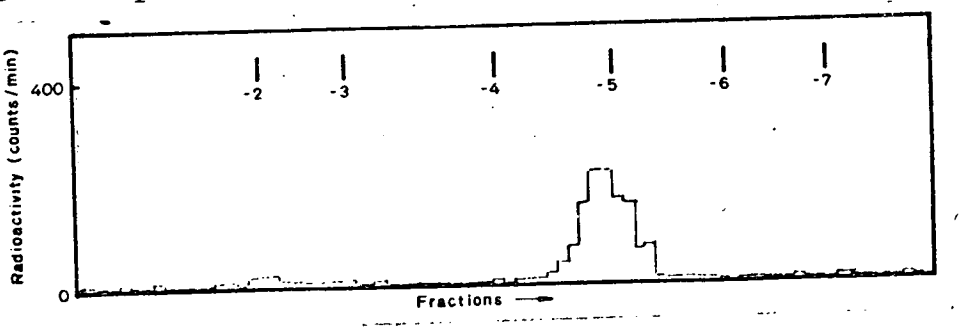


Fig.4.12. Rechromatography at pH 7.8 after a second round of alkaline hydrolysis.

Several unsuccessful attempts were made to label the tetranucleotide at the 5' end for sequencing purposes, using polynucleotide kinase and ^{32}P-ATP (see next chapter). This was the more annoying because the radioactive ATP contained an impurity with a charge of -5 which was for a while taken for pNmpNmpNmpN.

Eventually the tetranucleotide was sequenced using double labelled material. The desalted material was treated with alkaline phosphatase. The enzyme was removed by adding 100 µl of 0.15 M sodium acetate, 0.5% SDS, extracting with phenol/chloroform and desalting the aqueous phase once again. It was hoped to determine the 5' and 3' ends by fractionation of a total venom phosphodiesterase digest:

$$N_1mpN_2mpN_3mpN_4 \xrightarrow{\text{venom}} N_1m + pN_2m + pN_3m + pN_4$$

^{3}H	+	+	+	-
^{32}P	-	+	+	+

Fractionation of the digest by paper electrophoresis at pH 3.5 should give pN_4 as the one peak with no ^{3}H. To identify N_1m, a paper chromatography system was found which separated nucleosides but not 5' nucleotides (Fig.4.13). This was 3MM paper with a solvent consisting of 3:2:2:1 ethyl acetate, 7.5 M ammonia, isopropanol and n-butanol. The results of these two fractionations are shown in Figs. 4.14 and 4.15. From the paper chromatography the 5' end seemed to be G. The electrophoresis gave only one peak, under pA, with the ^{3}H and ^{32}P labels not quite coincident. So this is probably a mixture of pAm and pA, all three distal bases being adenine. This makes the complete structure of the tetranucleotide GmpAmpAmpAp.

The -4 peak was estimated to contain 7.8 phosphate

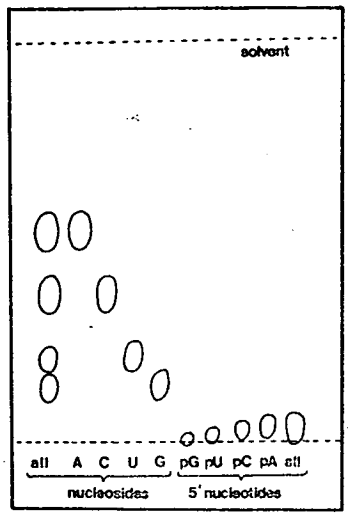


Fig.4.13. Paper chromatography system employed in Fig.4.14 to determine the 5' end of the tetranucleotide.

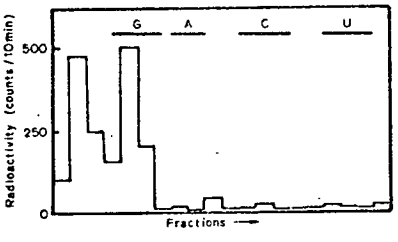


Fig.4.14. Fractionation of the total venom digest of the tetranucleotide by the chromatography system shown above.

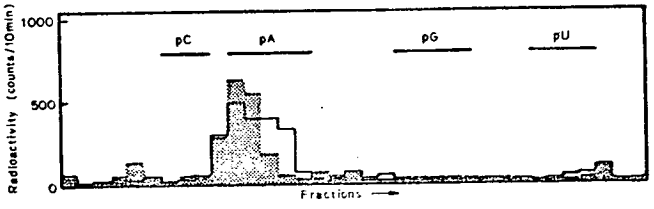


Fig.4.15. Paper electrophoresis of the total venom digest. The clear histogram represents ^{32}P radioactivity and the hatched one ^{3}H radioactivity.

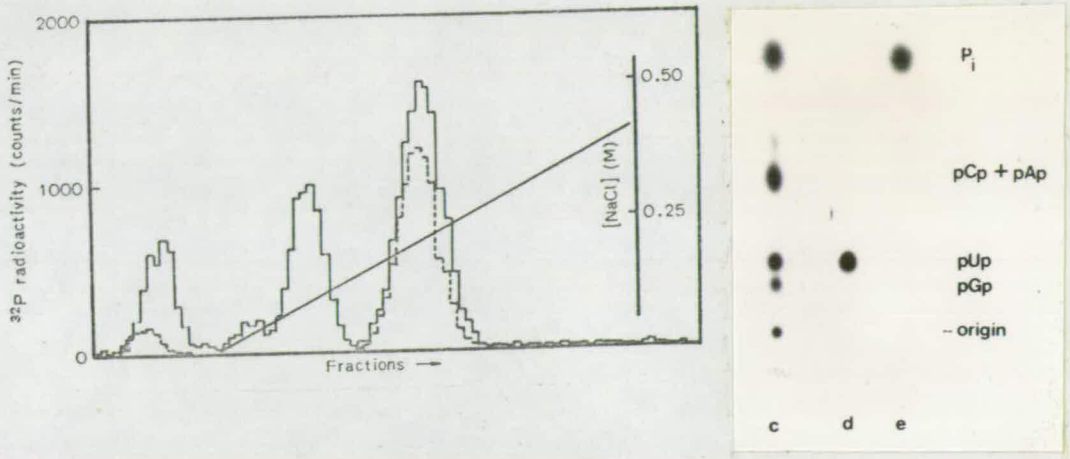


Fig.4.16. Resolution of the 28s -4 peak on DE-Sephadex at pH 2.7. The solid histogram represents ^{32}P radioactivity and the dashed histogram ^{3}H radioactivity.

Fig.4.17. PEI-cellulose chromatography. c) standards, d) peak -4/II from the 28s hydrolysate, e) the same after phosphatase.

groups per mole. It labelled with methionine to an extent representing 0.55 Me/P relative to the dinucleotides and 0.66 Me/P relative to the tetranucleotide. This was clearly a mixture of components and so a further resolution was attempted on DE-Sephadex at pH 2.7. This gave three peaks reproducibly, the first of which was a runoff peak. The relative areas of the three peaks was rather variable and the overall recovery from the column was about 60%. The methyl labelling pattern of the three peaks is shown in Fig.4.16. Both the first and the third peaks label and were assumed to consist of or to contain methylated trinucleotides. The position of elution from the column and the result of a total digestion with venom phosphodiesterase suggested that the third one was Gm(G,U). The first seemed to be a complex mixture on PEI-cellulose chromatography.

The second, unmethylated, component was assumed to

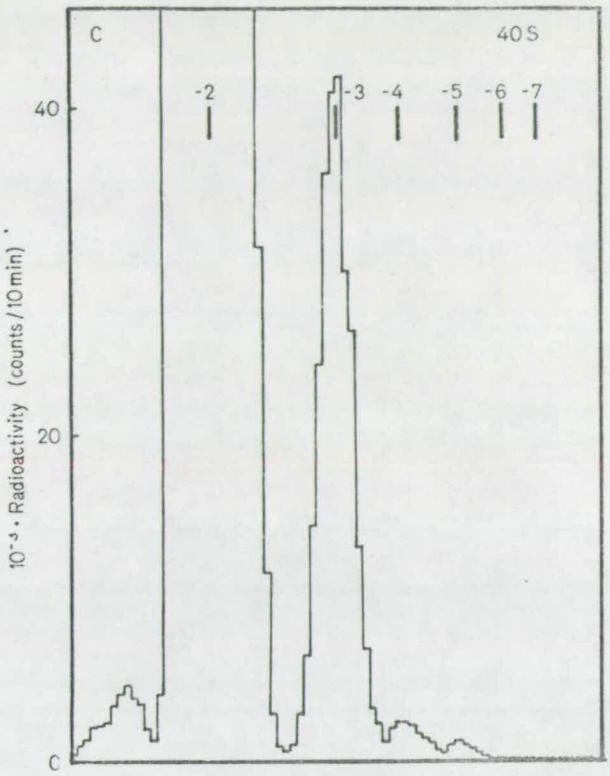
be the nucleoside-3',5'-diphosphate from the 5' end. As was the case for the 18s RNA, this co-eluted from the pH 2.7 column along with ATP, co-chromatographed on PEI-cellulose with pUp, and lost all its phosphate when treated with alkaline phosphatase (Fig.4.17). It was concluded that the 5' end of the 28s RNA is also pU.

40s RNA

The fractionation pattern of the alkaline hydrolysate is shown in Fig.4.18. Like the 28s pattern, this showed a relatively large -3 peak, and small -4 and -5 peaks. On no occasion was any peak seen above the -5 position, so it was decided that the precursor RNA does not bear a 5' triphosphate. It was not possible to label the 40s RNA to a high enough specific activity with (^{3}H-methyl)-methionine to enable the experiments involving this label to be carried out. So only the characterizations involving the ^{32}P label were done.

The -5 peak represented 4.5 phosphate groups per mole, and it seemed pretty likely that this was the same methylated tetranucleotide found in the 28s hydrolysate. Just to make sure, the shift out experiment was repeated with material from the 40s -5 peak and it gave the same result, viz peaks at -2 and -3.

The -4 peak was fractionated on a DE-Sephadex column at pH 2.7, and like the 28s material, gave three peaks. The first and third resembled those from the 28s RNA, while the second, which was the 5' end derivitive co-chromatographed with pGp on PEI-cellulose and lost all its phosphate on treatment with alkaline phosphatase (Fig.4.19). So the 5' end of the 40s RNA seems to differ from that of the 18s and the 28s and to be pG.



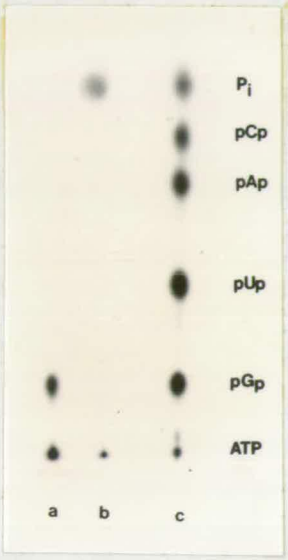


Fig.4.18. DE-Sephadex chromatography of 40s RNA hydrolysate. The bars show the isostichs and their charges and the histogram shows ^{32}P radioactivity.

Fig.4.19. PEI-cellulose chromatography. a) peak -4/II from the 40s hydrolysate, b) the same after phosphatase, c) standards.

Conclusion

Certain clearcut results may be summarized from these experiments:

i) The 5' ends of 18, 28 and 40s RNA are pU, pU and pG respectively.

ii) No rRNA species bears a 5' triphosphate (or diphosphate) such as may be found in total RNA or 5s RNA .

iii) The 40 and 28s RNAs contain a methylated sequence believed to be GmAmAmAp in addition to many isolated and one or two double ribose methylations.

These results will be discussed at the end of the next chapter.

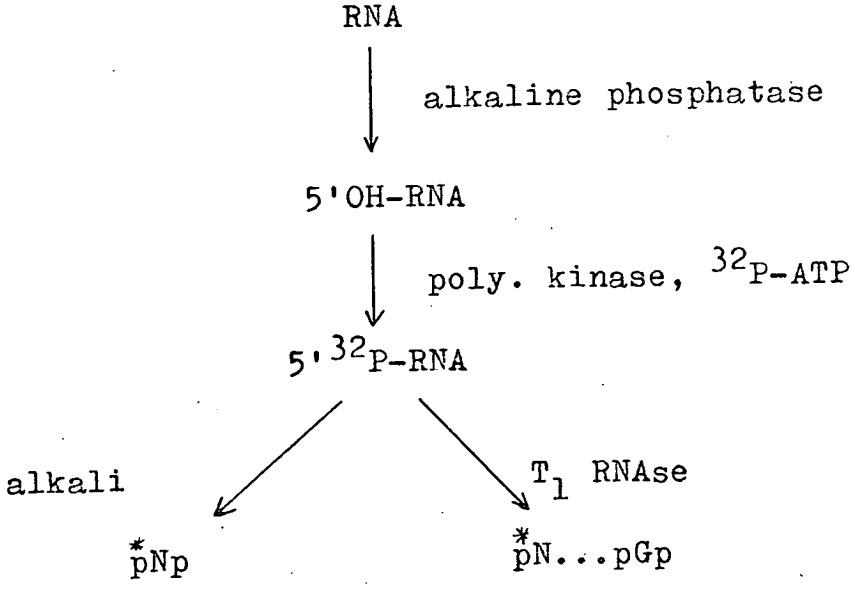
Chapter 5

INVESTIGATION OF 5' ENDS USING POLYNUCLEOTIDE KINASE

Introduction

This work was commenced before that described in the previous chapter. Because the results are consistent, but not so clearcut on their own, they are described second. The objectives of these experiments were the same.

In vitro terminal labelling of RNA with a radioactive phosphate group is an alternative method of 5' end determination to that described in the previous chapter. Alkaline hydrolysis of 5' labelled RNA yields a pNp as the only radioactive product, and digestion with a specific ribonuclease, *eg* T_1 RNAse, yields a single radioactive oligonucleotide which bears a phosphate group at each end. This method has been used to determine the terminal sequences of several bacterial rRNAs (116,117) and viral RNAs (118,119,120). In outline the procedure is as follows:



Labelling and Analytical Procedures

RNA Preparation

This was dealt with in chapter 2. Sterile precautions were especially important for these preparations, because the nicks in the RNA sequence can be labelled in the reaction, and may in fact be better substrates than the 5' end itself.

Removal of 5' Phosphate Group

The enzyme used was alkaline phosphatase from E.coli (Whatman biochemicals). Early trials with the Worthington enzyme showed the presence of considerable ribonuclease activity which could not be removed even by gel electrophoresis. The buffer used for phosphatasing was 50 mM Tris.Cl pH 7.6, 10 mM $MgCl_2$. The correct E:S ratio was arrived at by the following experiment. A sample of RNA was phosphatased and kinased and was dialysed exhaustively against 0.15 M sodium acetate, 0.5% SDS to remove radioactive ATP. Then samples were incubated for 30' at 37^{o} with different amounts of phosphatase. 1 ml of cold carrier RNA was added to each and 5 mls of 5% TCA. The samples were centrifuged and the supernatants counted by Cerenkov radiation and the precipitates counted on filters (toluene: 0.5% PBD, gain 1%, full window). The result is shown in Fig.5.1, and it can be seen that the reaction goes to its maximum extent with an E:S ratio of 1:2.

For the experiments to be described, unlabelled RNA was phosphatased under these conditions, the enzyme was removed by phenol extraction and the RNA precipitated twice with two volumes of ethanol.

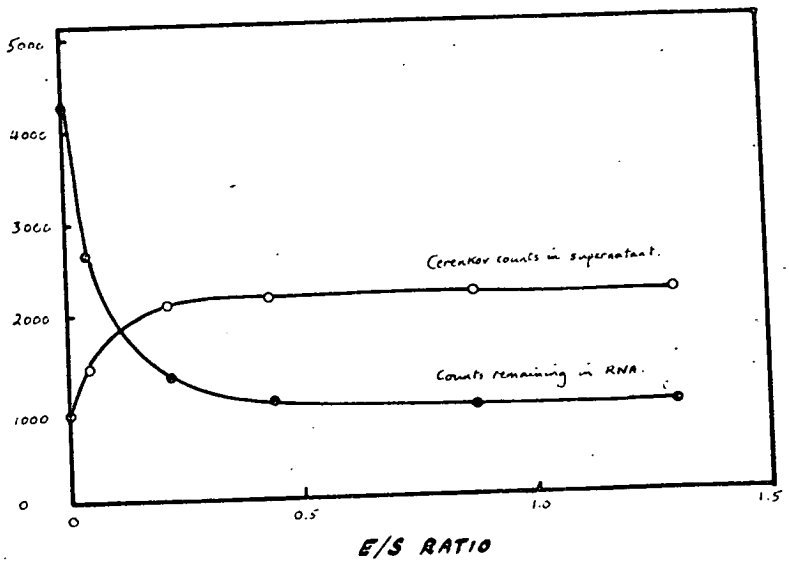


Fig.5.1. Efficiency of phosphatasing at different E:S ratios. Each pair of points represents a single tube, the open circles being the released phosphate and the closed circles being the retained phosphate.

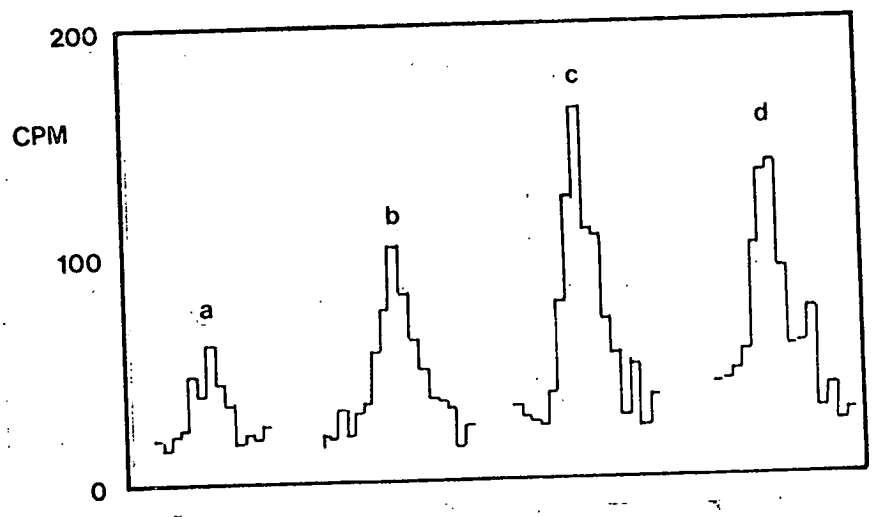


Fig.5.2. Efficiency of kinasing at different E:S ratios. Ratios in units enzyme per µg RNA are a) 0.1, b) 0.2, c) 0.4, d) 1.0.

5' Terminal Labelling

When used with DNA, this reaction apparently goes very readily under a variety of conditions. However, the present work with RNA indicated that the reaction conditions were very critical for success. The buffer used was 10 mM Tris Cl pH 7.4, 5 mM $MgCl_2$, 20 mM mercaptoethanol. It proved to be harder to decide on the correct E:S ratio than it was for the phosphatase. If there is a great excess of enzyme, then the amount of label incorporated still increases, because contaminating RNAse introduces nicks which become labelled. The way around this was to melt the samples in formamide and to separate the intact molecules from the degradation products by gel electrophoresis. Such an experiment is shown in Fig.5.2. Here the RNA was 28s and it seemed that the optimum E:S ratio was 0.5 units/µg. The incubation was for 45' at 37°.

The ATP came from Amersham and had a specific activity of 15 mCi/µmole. It was dissolved in water at 1 nmole/µl and used at 1/10th of this concentration, which is 5x the Km for the enzyme. So the reaction mixture was:

5 µg 5'OH-RNA
3 units PKase
2 µl γ^{32}P-ATP
buffer to 20 µl.

} incubation 37°, 45'.

Isolation of Product

Since the radioactive ATP was present in excess, it was necessary to remove it from the terminally labelled RNA. Gel electrophoresis proved to be an ideal technique for this, because the small and highly charged ATP molecules move quickly through the gel and into the

bottom buffer compartment. Also, any labelled degradation products of the RNA move faster than the complete molecules because they are smaller. The earlier experiments showed considerable (40%) heterogeneity for the 18 and 28s molecules. This was found to be caused by labelled nicks for it was abolished when the RNA was melted with formamide before loading onto the gel. In some experiments, whole cell RNA labelled with ^{3}H was added as a marker. This was to check that the ^{32}P labelled RNA was really of the same molecular weight as the original material, and had not been degraded. The RNA from the peak slices was eluted in lithium acetate and pelleted as described in chapter 2.

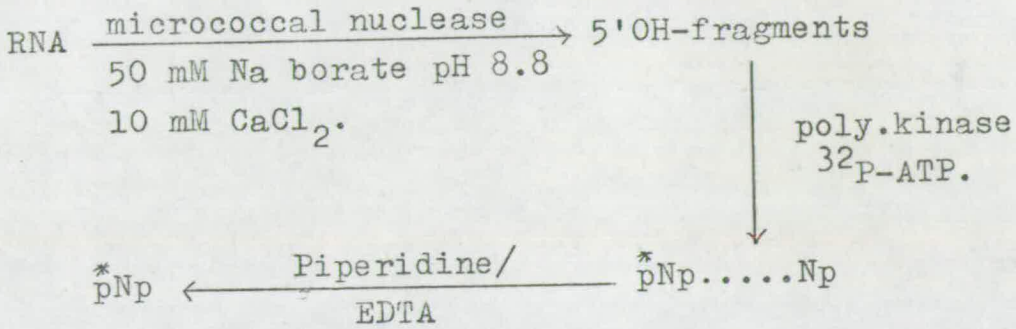
Alkaline Hydrolysis

Hydrolysis in KOH was not very successful for this purpose. Not all the salt can be removed by perchloric acid precipitation, and the remainder interfered with the chromatography of the nucleoside-3',5'-diphosphates. Success was obtained using 5% piperidine, 1mM EDTA, which is a mixture used to hydrolyse RNA from gel slices (14). The RNA was incubated in this at 60^{o} for 48 hours in a sealed capillary, and then the mixture was dried down onto a polythene sheet. The residue was redissolved in distilled water and dried down again.

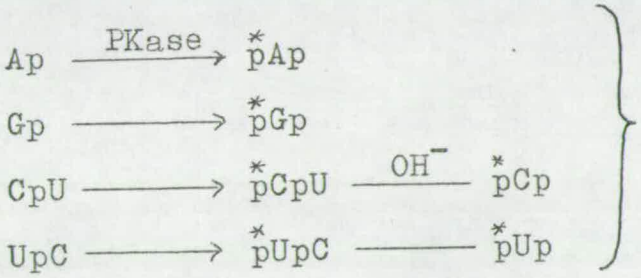
Chromatography of Nucleoside-3',5'-diphosphates

Before this work was carried out, there was no one-dimensional separation method for the four nucleoside-3',5'diphosphates. At first, paper electrophoresis was attempted, as this is the standard technique for the separation of nucleotides. But it was not satisfactory. Then PEI-cellulose thin layers were tried with various concentrations of pyridine formate pH 3.5. A concentration of 1 M seemed to give a satisfactory

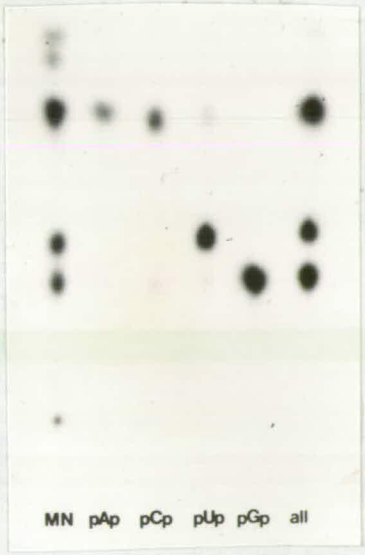
separation of the four compounds. Since it was not at first realized that the kinase would label 3' mononucleotides, the standards were prepared by the following reactions:



After a considerable time lag, the four radioactive compounds were prepared separately. The different routes of synthesis simply reflected the available starting materials:



When these became available it was clear that the



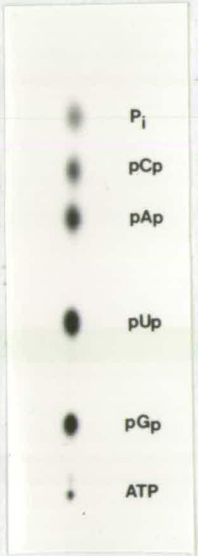


Fig.5.3. (left) chromatography of pNps in 1 M pyridine formate pH 3.5.

Fig.5.4. (right) chromatography of pNps in 2 M sodium formate pH 3.4.

system in use did not separate pAp from pCp, and that the leading spot from the micrococcal nuclease standards was some unidentified mystery compound. In fact this did not matter since none of the RNA species turned out to have pA or pC at their 5' ends, but a search for a suitable solvent led eventually to 2 M sodium formate pH 3.4 (Figs.5.3 & 5.4).

T_1 RNAse Digests

RNA pellets from the separation step were dissolved in distilled water and freed from salt and SDS by passage through 1 ml columns of G-100 Sephadex. The digestion was carried out on a polythene sheet as previously described. The sample was mixed with carrier RNA and enzyme and incubated at 37^{o} for 30' (E:S ratio 1:10, the buffer was 10 mM Tris Cl pH 7.4, 1 mM EDTA). Since the radioactive product has phosphate groups at both ends, it was displayed by running the mixture on a 2-dimensional separation technique used for pyrimidine tract analysis (110). This was PEI-cellulose with 2 M pyridine formate pH 3.5 in the first dimension and 1 M LiCl in the second dimension.

Elution of Material from Thin Layers

This was necessary for the preparation of standards and for the further analysis of the T_1 oligonucleotides. The piece of thin layer was cut out and the adsorbed material was chromatographed onto a paper flag with 30% TEC. Then it was eluted from the paper into a glass capillary with water.

Results

When all the techniques described above had been developed or got to work, it was in principle possible

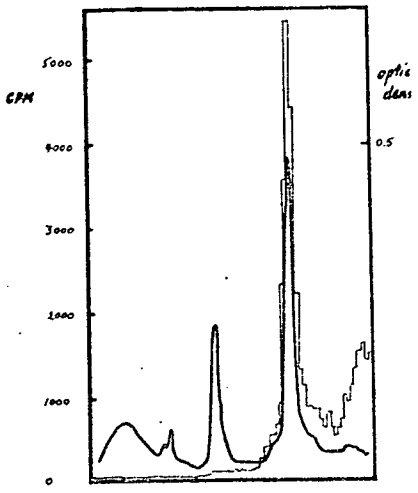


Fig.5.5.A Polyacrylamide gel of a mixture of unlabelled total RNA and 18s RNA labelled <u>in vitro</u> with polynucleotide kinase.

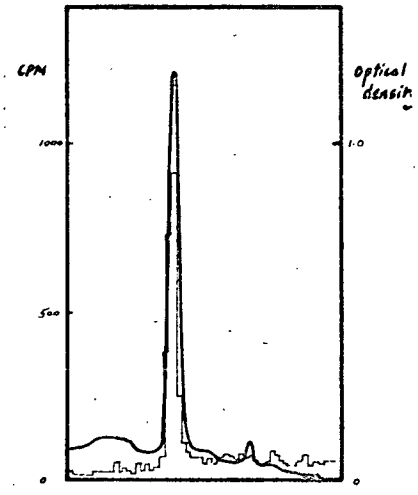


Fig.5.5.B The same for 28s RNA

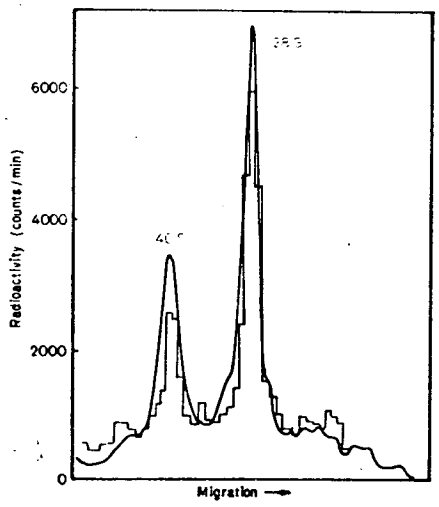


Fig.5.5.C Polyacrylamide gel of RNA from the first zonal rotor peak which has been phosphatased and kinased. The left hand peak is 40s and the right hand peak is 28s RNA.

to identify the 5' end of any RNA species. In fact the experiments are very tricky for one person because the RNA, the kinase, and the radioactive ATP are all unstable and have to be obtained at the same time for success. In addition the kinase reaction frequently failed to work for reasons which are not understood.

However, when the procedure did work, quite good labelling could be obtained (Fig.5.5). When tritiated RNA was run on the gels together with the in vitro labelled material, the isotope peaks were coincident, showing that the bulk of the sample had escaped degradation (Fig.5.6). In fact there must have been some nicking because the first end groups found for 18 and 28s RNAs were only 60-65% of pUp (Fig.5.7). When the labelled RNA was mixed with formamide before running

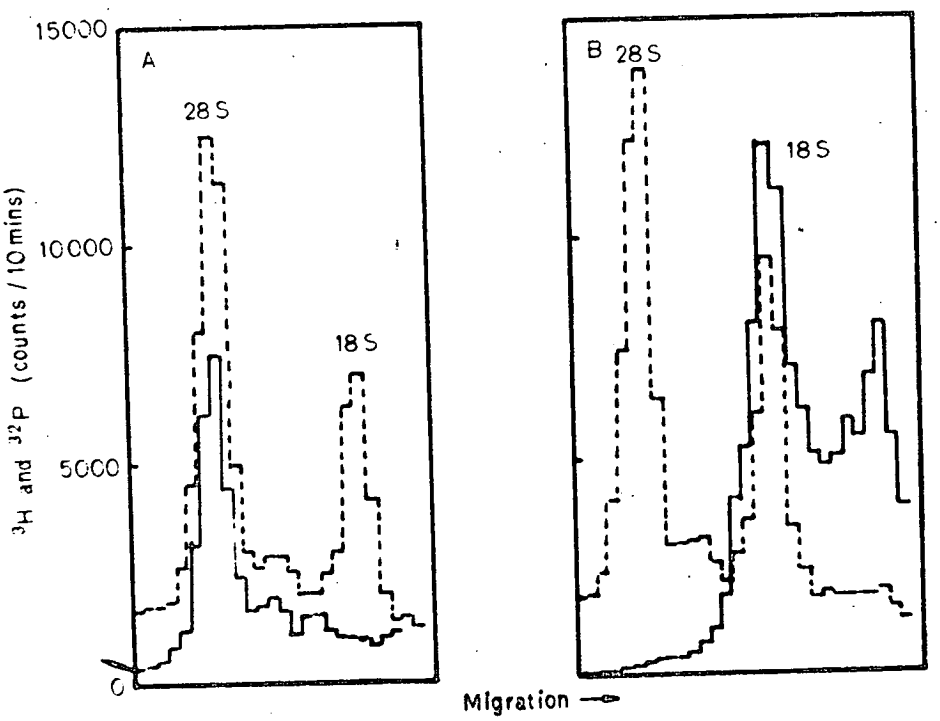


Fig.5.6. Double label experiment. This shows polyacrylamide gels of mixtures of 5' labelled species and total RNA labelled in vivo with ^{3}H-UR. The dashed profiles are the ^{3}H counts and the solid ones are the ^{32}P counts. A. 28s RNA and B. 18s RNA.

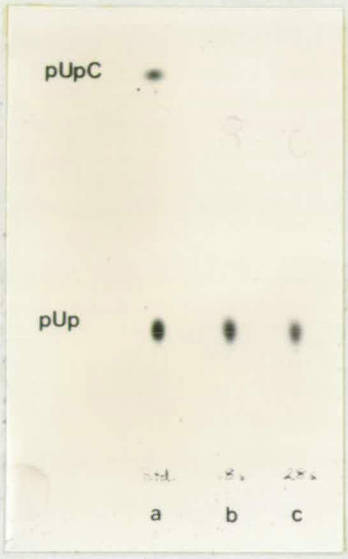


Fig.5.7. PEI-cellulose chromatography of pNps recovered from terminally labelled RNA species. a) 18s RNA, b) 28s RNA, c) standards. About 60% of the counts are in pUp for the two RNAs.
Fig.5.8. The same, except that the labelled RNAs were melted in formamide before gel electrophoresis. a) standards, b) 18s RNA, c) 28s RNA. 90% of the counts are in pUp for both RNA species.

the gel, the proportion of pUp recovered was more than 90% (Fig.5.8).

No T_1 product was obtained from the 18s RNA and no end group was successfully determined for the 40s RNA. But a comparison of T_1 products from the 28s and 40s RNAs was made, both with and without the formamide melting step. The predominant 40s product moved just behind the Orange G marker in the first solvent (Fig.5.9). This is the position of pGp, and when the end group had been determined by the alkaline hydrolysis method, it became clear that this was what it was, not an oligonucleotide at all.

The main product from the 28s RNA had about half this mobility in the first solvent. We know that its

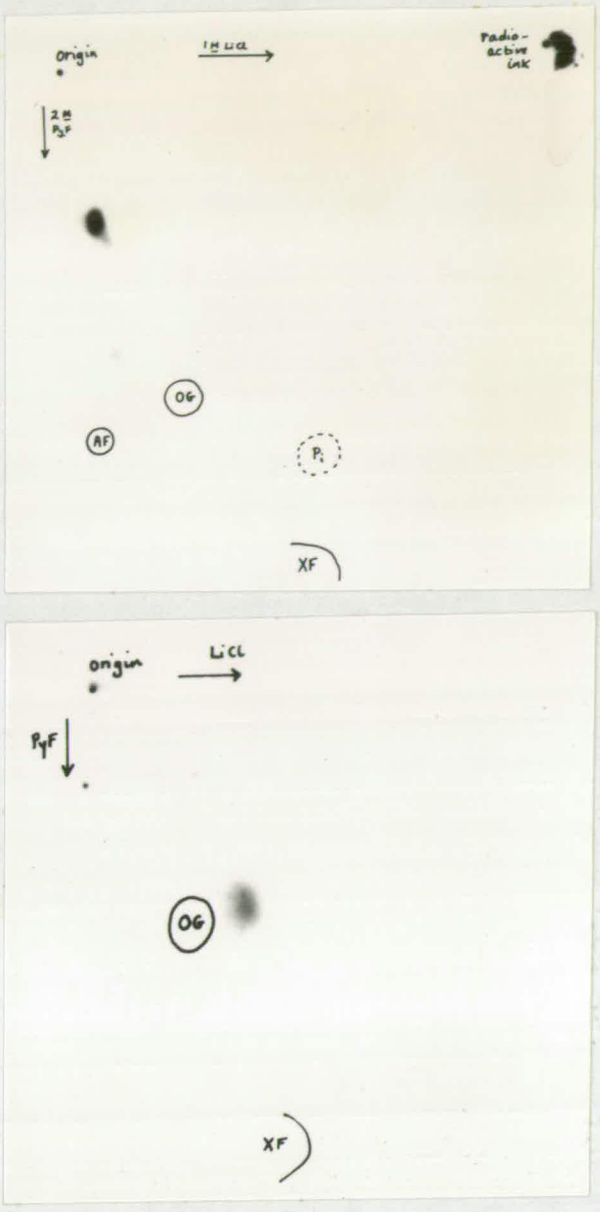


Fig.5.9. 2-dimensional chromatography on PEI-cellulose. The 1st dimension was 2 M pyridine formate pH 3.5, and the second was 1 M LiCl. The samples were T_1 digests of 5' labelled RNA species. The top plate is for 28s RNA and the bottom one for 40s RNA and it can be seen that the two spots run in different positions relative to the marker dyes.

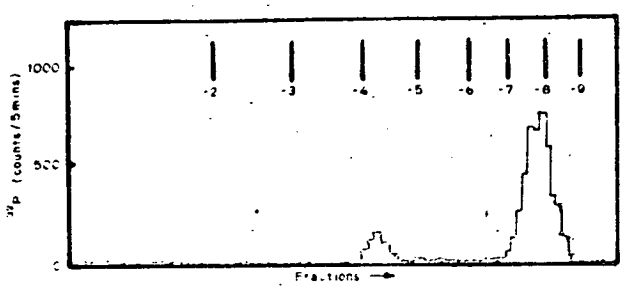


Fig.5.10. DE-Sephadex chromatography of 5' oligonucleotide from 28s RNA. The histogram shows the radioactivity and the bars show the positions of elution of T_1 isostichs made from yeast RNA.

5' end is pU and that its 3' end must be Gp because of the T_1 digestion. Its length was measured by DEAE Sephadex chromatography with unlabelled T_1 isostichs as markers. This column was run in exactly the same way as those described in the previous chapter (Fig.5.10). The oligonucleotide eluted with the seventh isostich and so must have a net charge of -8. The two terminal phosphates will bear a total charge of -4 and so there must be four internal groups each with a charge of -1. This makes the structure of the oligonucleotide pUpNpNpNpGp.

Discussion of Results from Chapters 4 & 5

Absence of 5' Triphosphate from 40s RNA

By analogy with 18 and 28s RNAs this finding suggests that the 40s RNA is itself a processed product, and that a larger RNA is transcribed, perhaps from the entire ribosomal repeating unit. In the author's opinion, there are several pieces of evidence against this. First, no giant rRNAs (mol.wt. 4 million) have been found in Xenopus despite a careful search in connection with the mechanism of nucleolar amplification (121). Secondly, the electron micrographs of ribosomal genes

in action (15) show an alternation of transcribed and non-transcribed regions. Thirdly, in the laevis-mulleri comparison (24), the spacer DNA seems to be more diverged than the excess, suggesting a greater selective pressure on the transcribed portion. On the other hand, evidence for the occasional transcription of spacer DNA comes from the observation of "prelude pieces" of RNA in the electron micrographs (122), and from the creation of polymers of 45s RNA by isolated rat nucleoli incubated in high concentrations of nucleoside triphosphates (123). If these examples can be regarded as pathological distortions of the normal process, then perhaps it may be suggested that the normal mode of progression of polymerase along the spacer involves going through all of the motions of transcription except the formation of the phosphodiester bonds. This would explain the 5' monophosphate and also the exceptional sensitivity of ribosomal RNA transcription to actinomycin D. Such a mechanism could be tested by showing that RNA polymerase I plus ribosomal DNA converts 5' triphosphates into 5' monophosphates. A plausible analogy may be the K restriction system of E.coli which has a large ATP requirement believed to be related to the process of translocation along the DNA duplex (77).

The End Groups

For the 18 and 28s RNAs, the occurrence of a 5' phosphate allows us to deduce that, whatever other properties the processing enzymes may have, they are nucleases of the B type. That is, enzymes which catalyse a direct hydrolysis by H_2O rather than phosphate transfers through a cyclic intermediate (124). Of course, such 5' phosphates have been found before in other eukaryotes eg L cells (106) and Saccharomyces cerevisiae (117), but this point does not seem to have

been realized before and it means, for example, that the methylations cannot protect sequences against processing, as suggested by Maden (51), because methylated sequences are vulnerable to the B type nucleases. In fact one such enzyme is the venom phosphodiesterase which is used in their analysis.

The fact that the 40s RNA has a single end group rather than a mixture is in accordance with the results described in chapter 3, ie that this substance is a unique molecular species and is not polydisperse. Although a study of the 3' end was not carried out, the author would be very surprised if it showed anything different.

The fact that the end groups of 28 and 40s RNA are not the same indicates that there is a segment of excess RNA at the proximal end of the precursor. Dawid's electron microscopic map of the precursor puts the 28s sequence at the extreme 5' end, but fortunatly electron microscopy is not yet sensitive enough to contradict the result described here. Both this finding, and the placing of the 5.8s sequence between the 18 and 28s (J.Speirs, personal communication) show that the processing mechanism must be more complicated than that proposed by Perry and Kelley (72), which envisaged a single endonucleolytic cleavage followed by 3' exonuclease digestion.

The Trimethylated Sequence

At first the discovery of the trimethylated sequence was rather surprising because no such sequence had been found before, even in an exhaustive analysis of the HeLa rRNA methylations (58). But a telephone call to B.E.H.Maden revealed that he suspected the presence of such a sequence in HeLa 28s RNA and believed

that the alkali stable fragment might have been mistaken for the 5' end fragment pCmpUp by Choi and Busch (81). This latter result was of course much quoted as evidence that the 28s sequence lay at the extreme 5' end of the precursor.

Somewhat later, when some rat cells were being grown in the laboratory the opportunity was taken to examine this possibility. Rat 28s RNA was prepared and the alkaline hydrolysate fractionated on a DEAE column. The -5 peak was present. It was shown to be labelled by (^{3}H-methyl)S-adenosyl methionine in isolated nucleoli and it yielded a species of net charge -3 after phosphatase treatment. So it did seem to be a methylated tetranucleotide similar to that in Xenopus.

Recently, Nazar and Busch (125) have retracted the original claim and now propose that the -5 species from rat is the methylated tetranucleotide AmGmCmAp. Although the sequence is different, this result is much more in accordance with ours, and it is always possible that there is a genuine difference of sequence between rat and Xenopus.

The function of the ribose methylations remains unknown. It seems probable that they do have a function because they are restricted to the 28 and 18s segments of the precursor and because they are all quantitative. In addition, the probability of obtaining three adjacent methylations from randomly scattered groups is about 0.5×10^{-3}. However, the only experimental evidence which bears on the problem is that of Vaughan et.al.(72) who found that processing was arrested at the 32s stage in methionine starved cells, and that this RNA was undermethylated. This at least showed that the methylations are not necessary for the specificity of the early processing reactions.

Chapter 6

A STATISTICAL TREATMENT OF RNA DIGESTS

Introduction

The work on end groups described in the preceding chapters had acquainted the author with some of the techniques used in RNA sequence analysis. It was thought probable that information other than the sequence could rapidly be obtained from the data provided by the oligonucleotide separations, if there existed a statistical theory of digests. At present, the field of primary sequence analysis is a very qualitative one, but a quantitative approach will eventually become a necessity, especially for sequence studies of DNA. It was with the conviction that a quantitative theory could be created that the work described below was undertaken.

The digests which are considered are those of RNA by RNAse A (cuts after pyrimidine nucleotides) and RNAse T_1 (cuts after G residues), and the separation techniques which are given special attention are the isostich fractionation on DEAE columns (103) and the "fingerprinting" technique of Sanger, Brownlee and Barrell (126).

The following terms will be used repeatedly in the next three chapters: isostichs are sets of oligonucleotides of common chain length; compositional isomers are sets of common length and base composition; and sequence isomers consist of defined sequences of nucleotide residues. The term instance will be used to refer to each copy of an isostich, compositional isomer or sequence isomer; and the incidence of a given species is the number of instances in the entire sequence.

The total sequence complexity will be defined according to the usage of Wetmur and Davidson (127) as the total unique sequence length measured in nucleotide residues. Note that for a defined species of RNA such as a single messenger RNA the complexity is the same as the length of the molecule, while for in vitro transcripts from DNA it may be quite different.

An ideal sequence will be defined as one in which the four nucleotide residues occur with equal frequency and the probability of a particular residue in a particular position is independent of the sequence on either side. A random sequence will be defined more loosely as one which was or could have been created by a stochastic process, but which may include constraints which make it non-ideal. Some of the following theory is derived for ideal sequences only, and some covers random sequences which shows some straightforward type of non-ideality such as a particular base composition.

The extent to which the "ideal nucleic acid laws" will cover real nucleic acids will be considered in chapters 7 and 8 in the context of their application to ribosomal RNA and to the properties of multiple gene arrays in general.

Number and Weight Distributions for Isostichs

The use of DEAE columns with 7 M urea to separate the isostichs of a nucleic acid digest has prompted some quantitative treatment of the distribution of ^{32}P radioactivity between the isostichs (128-130). Of these references, the last is the most detailed but it is in error in its formula for the number fraction of each isostich because the break points are each counted

twice: at the beginning and end of each run.

For the purposes of the following calculation, an RNA sequence will be regarded as a sequence of Bernouilli trials which are "successes" or "failures" depending on whether they are on the immediate 5' side of a chain scission. Then if we ask what proportion of the molecules in the digest are of length n, this is the same as asking what is the probability of a run of n-1 "successes" in the sequence. Such probabilities are given by the geometric distribution, so if the proportion of break points ("failures") is y:

$$\text{Proportion of molecules length } n = y(1-y)^{n-1}.$$

This is a number distribution. If we now wish to know the proportion of radioactivity in each isostich we must convert this to a weight distribution. This is done by multiplying each term by n and renormalizing so that the sum of all terms is unity. Normalization is carried out by dividing by the mean of the geometric distribution (1/y), so that the result is:

$$z(n) = ny^2(1-y)^{n-1} \quad \ldots\ldots (1)$$

where z(n) is the fraction of radioactivity in the n th isostich. It is convenient to put this equation into linear form by taking logarithms:

$$\log\left(\frac{z}{n}\right) = (n-1)\log(1-y) + 2\log y \quad \ldots (2)$$

Now if log(z/n) is plotted against n-1, a straight line is obtained with a slope of log(1-y) and an intercept of 2 log y.

The important thing to note about this relationship is that z is a function only of the proportion of break

points and not of the total complexity of the sequence. In other words, long runs do not occur in a larger proportion in long sequences. However, the equation is only exactly true for a sequence of infinite complexity. For finite sequences of complexity T there will be sampling errors in the actual weight fractions and these will be greater the smaller the value of T.

For this reason we need confidence limits for the points represented by equation (2). The statistical analysis of this situation has been worked out in terms of a probability distribution for the number of fragments of length n (D.Sales, unpublished).

The mean of this distribution, which is the expected value of the incidence of the n th isostich, is denoted by R:

$$R = \frac{Tz}{n} \quad \ldots\ldots (3)$$

and the standard deviation, which is the standard error of R is given by:

$$\sqrt{\frac{Tz}{n}\left(1 - \frac{z}{n}\right)}$$

If $\hat{z}$ is defined as a random variable with a mean specified by equation (1), then:

$$S.E.\left(\frac{\hat{z}}{n}\right) = \frac{1}{T}\sqrt{\frac{Tz}{n}\left(1 - \frac{z}{n}\right)} = \sqrt{\frac{z}{nT}\left(1 - \frac{z}{n}\right)}$$

and the standard error of the transformed variable $\log(\hat{z}/n)$ may be found by using a Taylor series approximation (131):

/over

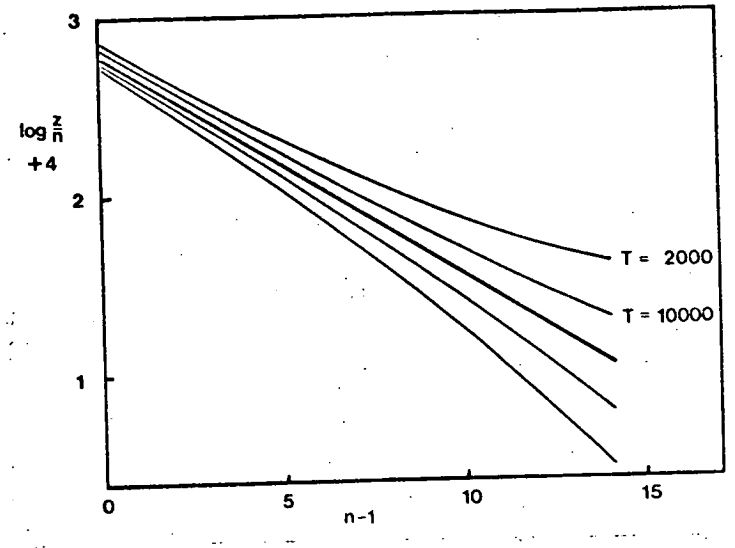


Fig.6.1. Linear plot of equation (2) with confidence limit envelopes for two values of T.

$$S.E.[f(y)] = f'(\mu_y) \cdot S.E.(y)$$

The result is:

$$S.E.\left[\log \frac{\hat{z}}{n}\right] = \log_{10} e \sqrt{\frac{1}{T}\left(\frac{n}{z} - 1\right)} \quad \ldots (5)$$

When this equation is used to calculate confidence limits around the straight lines represented by equation (2) we obtain a family of curves having T as parameter (Fig.6.1). Of course, these curves are not confidence limits for the whole straight line, but they join the 95% limits for each point considered independently.

The use of this linear plot to analyse isostich separations will be demonstrated in the next chapter.

Properties of RNA "Fingerprints"

Fingerprinting is a two-dimensional separation technique which resolves oligonucleotides mainly by

their length and base composition. The experimental details will be described later. For the present it should be noted that the arrangement of spots shown on the autoradiograms corresponds approximately to a catalogue of compositional isomers in the digest.

In the opinion of the author, the term "fingerprint" is rather misleading because it implies that the pattern of spots is uniquely characteristic for each RNA species. This may be true for the large and rare oligonucleotides but the smaller ones are likely to be present in all RNA sequences and this is why all "fingerprints" of long RNA sequences do look roughly the same. From a quantitative point of view, the question is: can we predict the properties (number and intensities of spots) of a fingerprint from the overall characteristics of the RNA (base composition and complexity) ?

Number of Spots

The number of spots approximates to the number of different compositional isomers present in the digest. This may appear to be a simple quantity to calculate but it is in fact quite a formidable combinatorial problem. There are several stages to the calculation: i) What is the maximum number of sequence and of compositional isomers which can be present in a particular isostich ? These quantities will be respectively designated by g(n) and f(n) and can be deduced from simple probability arguments. For example, the f(n) for T_1 oligonucleotides is the number of ways of partitioning n-1 objects into three groups. The results are:

/over.

		RNAse A	RNAse T_1
number of compositional isomers	$f(n)=$	$2n$	$\frac{n(n+1)}{2}$
number of sequence isomers	$g(n)=$	2^n	3^{n-1}

ii) What is the number of sequence isomers included in a given compositional isomer ? If the compositional isomer is written as $A_aC_bU_cG$, then the number of sequence isomers will be given by the multinomial coefficient:

$$\frac{(n-1)!}{a!\,b!\,c!} \qquad \text{where } a+b+c=n-1$$

iii) What is the expected incidence of the n th isostich in a sequence of complexity T ? This is given in equation (3) above:

$$R(n,T) = \frac{Tz}{n}$$

iv) If we consider the n th isostich, then the expected number of spots which it contributes to the fingerprint will be the expected number of the f(n) isomers which are represented among the R(n,T) instances. This would be a classical occupancy problem if the probabilities of the compositional isomers were all the same. But since they depend on the number of included sequence isomers we must adopt a procedure described by Feller (132).

If the compositional isomers are numbered from 1...i...f, then:

$$P(i) = \frac{(n-1)!}{a!\,b!\,c!} \times \frac{1}{g(n)}$$

and the probability that the i th compositional isomer will be unrepresented in the R instances of the isostich is then:

$$[1 - p(i)]^R$$

v) In order to find the probability that exactly j of the f isomers are not represented, we have to calculate the coefficients S_k. These are the sums of the probabilities of all the ways in which k compositional isomers can be absent from R instances:

$$S_K = \sum_{i_1=1}^{f} \sum_{i_2=i_1+1}^{f} \cdots \sum_{i_K=i_{K-1}+1}^{f} [1 - \sum^{K} p]^R$$

Then:

$$p(j) = S_j - {}^{j+1}C_j S_{j+1} + {}^{j+2}C_j S_{j+2} \cdots \pm {}^{f}C_j S_K$$

where the C terms are binomial coefficients. This is a probability distribution for the absence of j out of the f possible isomers which is the same as the probability that f-j are present in one or more copies. So the expected value for the number of spots contributed by the n th isostich is f(n) minus the mean of this distribution.

vi) For the whole fingerprint, these values would be summed over all the isostichs.

Unfortunately, there is no analytical expression for the mean of the probability distribution, and the calculation of the coefficients S_k would require enumeration of all the selections of k from f isomers. So although the spot number could be calculated by this method, it would be somewhat tedious to do so.

Of more practical use is an approximation which was discovered in the course of research on the spot number problem. It is a relationship between the spot number

per isostich, X, and the isostich length, n:

$$X(n,T) = 0.9\, f\left[1 - \left(1 - \frac{1}{f}\right)^{R}\right] \quad \ldots\ldots (6)$$

The total spot number is obtained by summation over the isostichs:

$$N(T) = \sum_{n=1}^{n=\infty} X(n,T) \quad \ldots\ldots\ldots\ldots (7)$$

This equation arises if it is assumed that all the compositional isomers have an equal probability of 1/f, and that their occurrence or non-occurrence in R trials are independent of one another. Then the right hand side of equation (6) can be seen as:

$$0.9 \times \begin{matrix}\text{number of}\\ \text{compositional}\\ \text{isomers}\end{matrix} \times \begin{matrix}\text{probability of}\\ \text{presence of a}\\ \text{given isomer.}\end{matrix}$$

The errors brought in by the two assumptions largely cancel, but the factor 0.9 is inserted to give better agreement with the computer simulations which will be described below.

If equation (6) is written out in full then for a T_1 digest:

$$X(n,T) = \frac{0.9\, n(n+1)}{2}\left[1 - \left\{1 - \frac{2}{n(n+1)}\right\}^{Ty^2(1-y)^{n-1}}\right]$$

Some plots of this equation are shown in Fig.6.2. For the shorter isostichs, all the possible compositional isomers are present, then as n increases the number present first rises to a maximum and then falls to zero. The proportion of the longer isomers is increased by a fall in y, but the most important feature of the relationship is that, unlike the z function, it is strongly dependent on the sequence complexity.

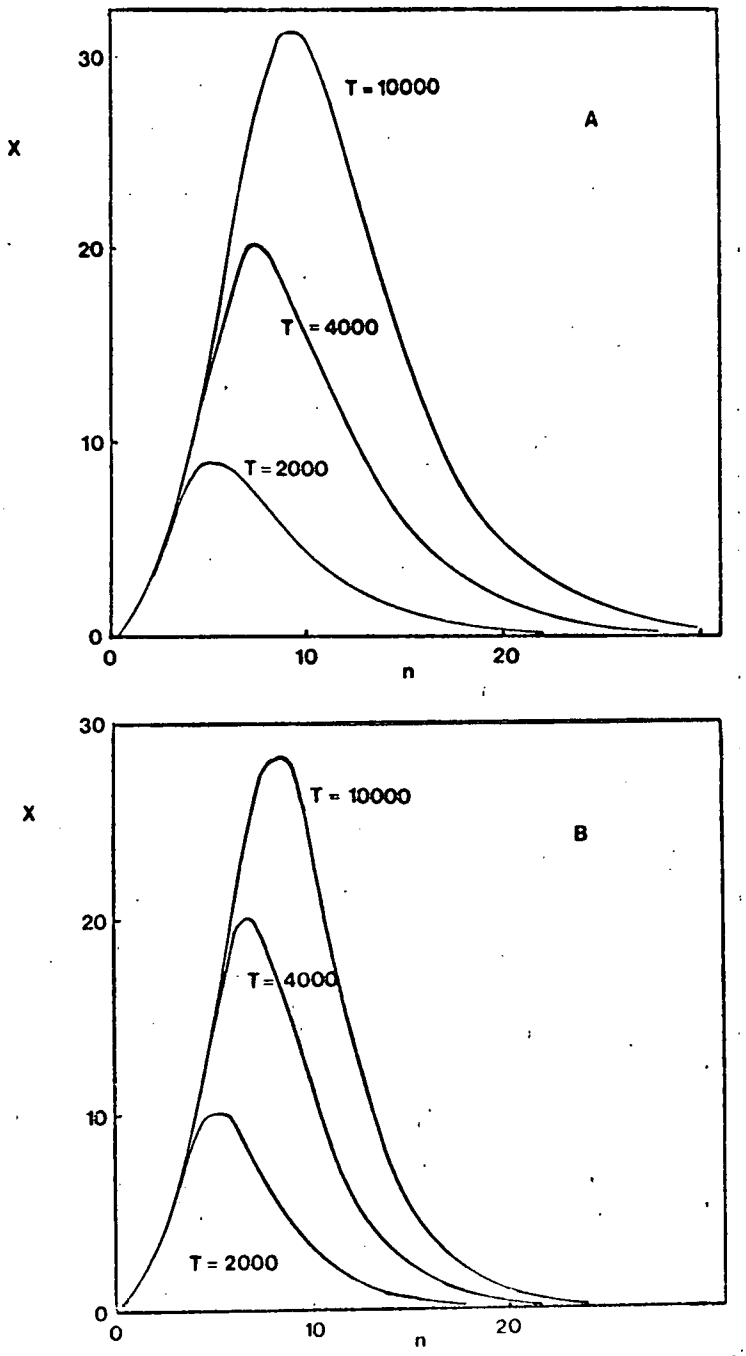


Fig.6.2. Plots of equation (6) (number of compositional isomers per isostich) for various values of T. A. y=0.2 B. y=0.3.

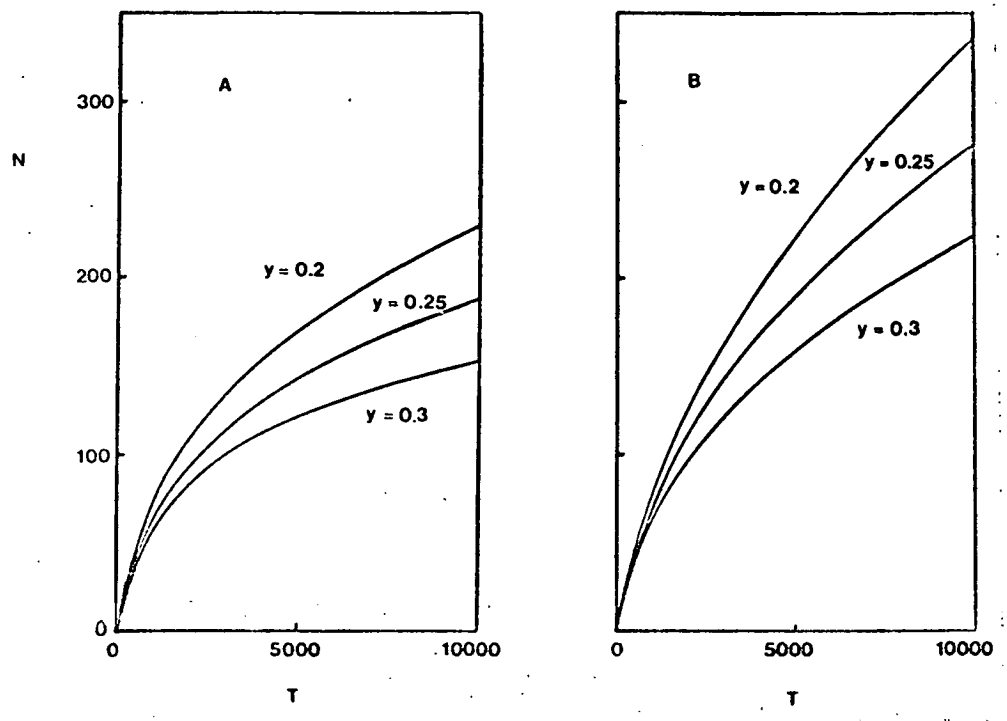


Fig.6.3. Plots of equation (7) (total number of compositional isomers v. complexity) for various values of y. A. RNAse A digest, B. RNAse T_1 digest.

In Fig.6.3 are shown plots of the total spot number against complexity for both RNAse A and RNAse T_1 digests. Especially for the latter there is quite a strong dependence on T, which makes it a practical possibility to estimate the complexity of a sequence from an experimental spot count.

It should be stressed that in addition to the small systematic errors resulting from the approximations there will be statistical fluctuations of experimental values of X(n,T) around the predicted ones. These arise first from the standard error in R, expressed in equation (4), and secondly from sampling errors involved in the representation of the different compositional isomers.

The Calculation of Intensities

The simplest way of representing the intensities of the spots on a fingerprint is by a histogram of spot frequency on intensity. This will be called an intensity spectrum and designated by a continuous frequency distribution $\emptyset(I)$. I represents intensity and will be normalized so that the sum of all intensities on a fingerprint is unity.

Again each isostich will be considered separately. For the n th isostich, the intensity of a single instance of any compositional isomer is n/T, and if the incidence is j, then the intensity is jn/T. The distribution of instances of the f(n) compositional isomers among many sequences will follow a multinomial distribution in which the probability of the i th compositional isomer is p_i. The number of trials is R(n,T), so:

$$P(k_1, k_2 \ldots k_f) = \frac{R(n,T)!}{k_1!\, k_2! \ldots\ldots k_f!} p_1^{k_1} \cdot p_2^{k_2} \ldots p_f^{k_f}$$

where k_i is the incidence of the i th isomer.

For a single isomer, the distribution will be a binomial in which each term will be a summation over all terms having k_i instances of the i th isomer:

$$P(k_i) = {}^{R}C_{k_i} \cdot p_i^{k_i} (1 - p_i)^{R-k_i}$$

To move from this equation to an expression for the intensity spectrum it is necessary to make two simplifications. First, as before, to assign each compositional isomer the same probability 1/f; and secondly to assume that the frequency distribution for a single isomer between sequences approximates

to the frequency distribution for the collection of isomers in a single sequence. The author knows of no mathematical justification for this second step, but it seems to work. Since there are X(n,T) compositional isomers in the n th isostich this frequency becomes:

$$W(j) = X(n,T)\,\frac{R(n,T)!}{(R(n,T)-j)!\,j!}\left(\frac{1}{f(n)}\right)^{j}\left(1-\frac{1}{f(n)}\right)^{R-j}$$

Substituting: $I = jn/T$

$$w(I,n,T) = \frac{X.R!}{\left(R-\frac{TI}{n}\right)!\left(\frac{TI}{n}\right)!}\left(\frac{1}{f}\right)^{\left(\frac{TI}{n}\right)}\left(1-\frac{1}{f}\right)^{\left(R-\frac{TI}{n}\right)}$$

and summing over the isostichs:

$$\phi(I,T) = \sum_{n=1}^{n=\infty} w(I,n,T) \quad \ldots\ldots(8)$$

In practice, the intensity spectra were calculated using the normal approximation where $R \geqslant 50$ and putting $\phi=0$ outside the range of the binomial.

Some spectra are shown in Fig.6.4. They are highly dependent on T with respect to the position of the maximum and the dispersion. The effect of base composition (the y parameter) mainly operates through the effect on N(T) which is the area under the curve (not shown).

Since the approximations involved in this calculation are so extreme, it is not claimed that the equations provide accurate quantitative predictions about the properties of actual fingerprints. But they do give a good qualitative indication of their behaviour, as will be illustrated in the next chapter. As before, we may expect statistical fluctuations in the experimental results to be superimposed on the systematic errors.

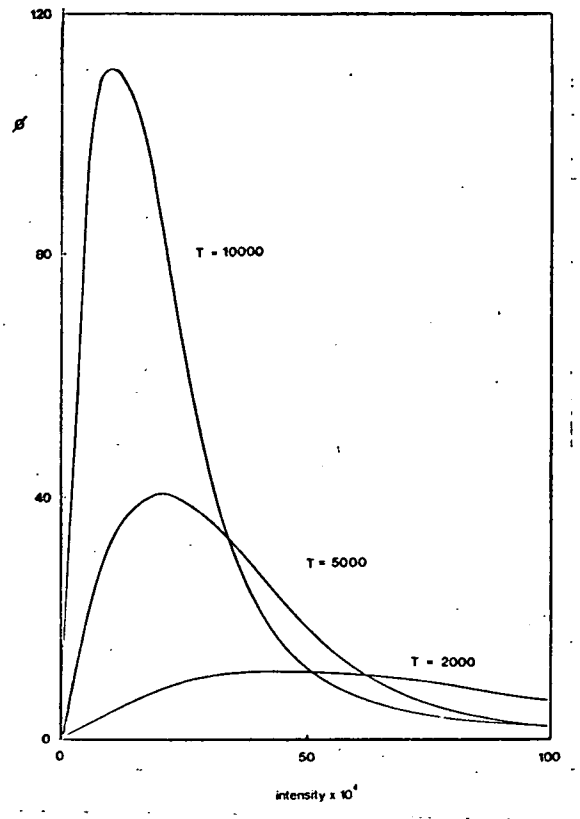


Fig.6.4. Intensity spectra for sequences of three different complexities. y=0.25.

Discussion

This theoretical work opens up various new possibilities. The log(z/n) vs. n-l plot could be used as a test of randomness in an RNA sequence and is used in this way in the next chapter. It can also be used to determine the proportion of "break points" in all analagous problems involving the geometric distribution. For example, when satellite DNAs are digested with restriction enzymes, a "ladder" of fragments is obtained if the repeating unit contains a restriction site. The log(z/n) vs. n-l plot has been applied by E.M.Southern and H.J.Cooke of the MRC Mammalian Genome Unit to measure the degree of divergence of various satellites. The "y" parameter given by the plots is the proportion of restriction sites which has survived the divergence, and if the site contains ν nucleotides:

/over.

probability of alteration of a single nucleotide $= \frac{(1-y)}{\nu}$

The existence of a formula relating spot number of fingerprints to sequence complexity should make it possible to measure complexities of mixed samples, eg total messenger RNA, or RNA transcripts of reiterated DNA, which are hard to measure in other ways. It is known for example that the measurement of complexity from reassociation rate is not reliable for reiterated DNA because of the effects of mismatching on reaction velocity (133,134).

The possible applications of intensity spectra will be considered in chapter 8.

Chapter 7

COMPUTER SIMULATIONS AND EXPERIMENTS ON RNA DIGESTS

Simulations

The expressions for z, X and $\emptyset$ in the previous chapter are based on the assumption of ideality in the RNA sequence. Furthermore, the working relations for $\emptyset$ and X are approximations. It is therefore necessary to ask two questions of the analysis:

i) Are the approximations valid, and over what range of the parameters do they hold ?

ii) How non-ideal must a sequence be before it ceases to obey the equations ?

These questions have been answered by means of computer simulations of RNA digestions. Simulations have several advantages over experiments, in particular the properties of the sequence are known exactly at the start and each parameter can be varied independently. Also, many simulations can be carried out in the time it takes to do one experiment, and there is no question of experimental errors.

The general nature of the computer programs used is described in appendix 2. The most important program was AJSSAT, which constructed digests from sequences supplied as data. The sequences used were random sequences of specified base composition composed by the program AJSSEQ. There is a certain problem about non-ideality because there is an infinite number of ways in which a sequence could be non-ideal. What was done for the simulations described below was to construct Markov chains from the four letters A,C,U,G which were characterized by a single parameter, r.

All the diagonal elements of the transition matrix were put equal to r and all the non-diagonal elements to (1-r)/3. The result of this is that sequences with $r > 0.25$ have a propensity to contain runs of the same base, while those with $r < 0.25$ have a deficiency of such runs.

z function

z(n) may be calculated from a simulated digestion by summing all the spot intensities for each isostich. The prediction from equation (2) is that we shall obtain the same straight line plot of log(z/n) vs. n-1 for sequences of the same base composition but different complexity. The scatter of points about the line should be less the greater is the sequence complexity and should be indicated by the confidence limits calculated from equation (5).

In Fig.7.1 are shown two plots from simulated digests. The straight lines are the theoretical plots for y=0.25, and the points are obtained from simulations of digestions of ideal sequences. In the upper figure, T=2000 and in the lower figure T = 10000, and it may be seen that the scatter differs accordingly. If the proportion of potential break points in the sequence is altered, then the slope and intercept of the plots should change accordingly, and this was found to be so.

In Fig.7.2 are shown plots for a series of Markov chain sequences having r= 0.1, 0.2, 0.3, 0.4. The deviations from the straight line are readily comprehensible: if there is a propensity for runs of identical bases then there will be more G in the digest relative to the other short isostichs, and vice versa. But these deviations seem to be small compared to the statistical scatter in the range $0.2 > r > 0.3$ and so we may consider

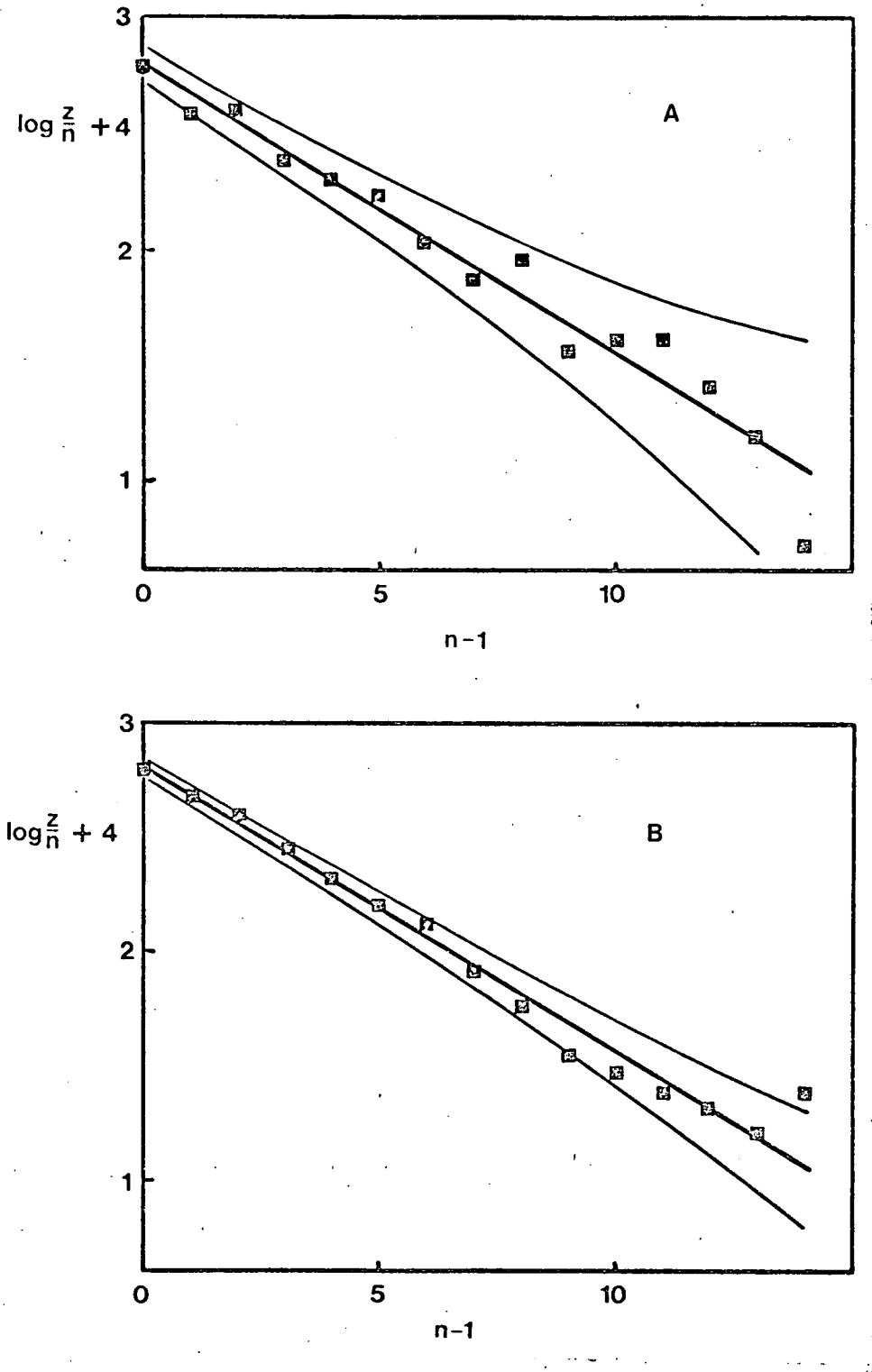


Fig.7.1. Linear plots calculated from simulations of RNA digestions. For both, the straight line is the theoretical plot for y=0.25, the curves are confidence limit envelopes for the appropriate complexity, and the points are from a simulation with a sequence of that complexity. A. T=2000, B. T=10000.

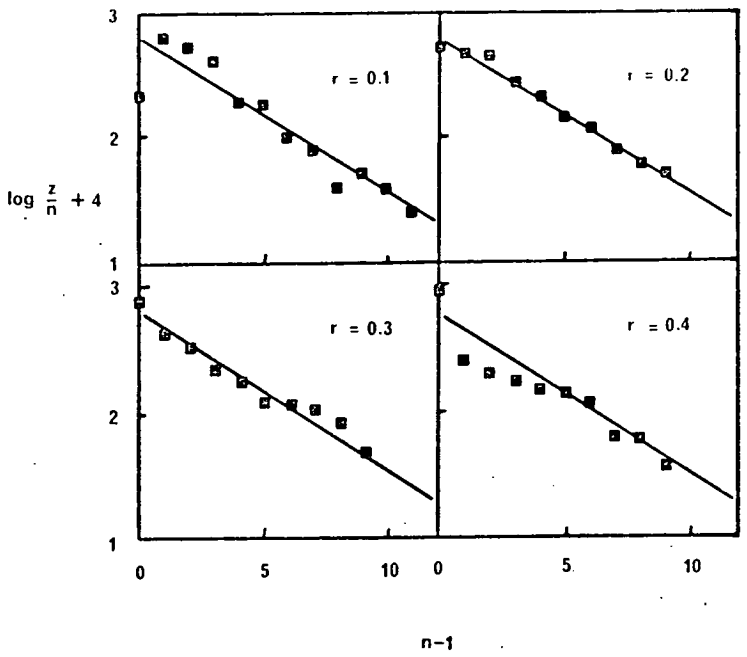


Fig.7.2. Linear plots for Markov chain sequences at four values of the r parameter.

the equation to be valid for such non-ideal sequences.

X function

X, the number of compositional isomers in each isostich of a digest, is given directly by the simulations (see appendix II). When plotted against the isostich length, this shows good agreement with the predictions of equation (6) as far as the effect of complexity is concerned (Fig.7.3).

Values of N, the total number of different compositional isomers in the digest, are also obtained directly from the simulations and are shown in Fig.7.4 for various values of T and y. Equation (7) seems to hold quite well below T=5000 but to overestimate the number of isomers at higher complexities.

For Markov chain sequences (not shown), the total spot number shows a steady increase with r, but the

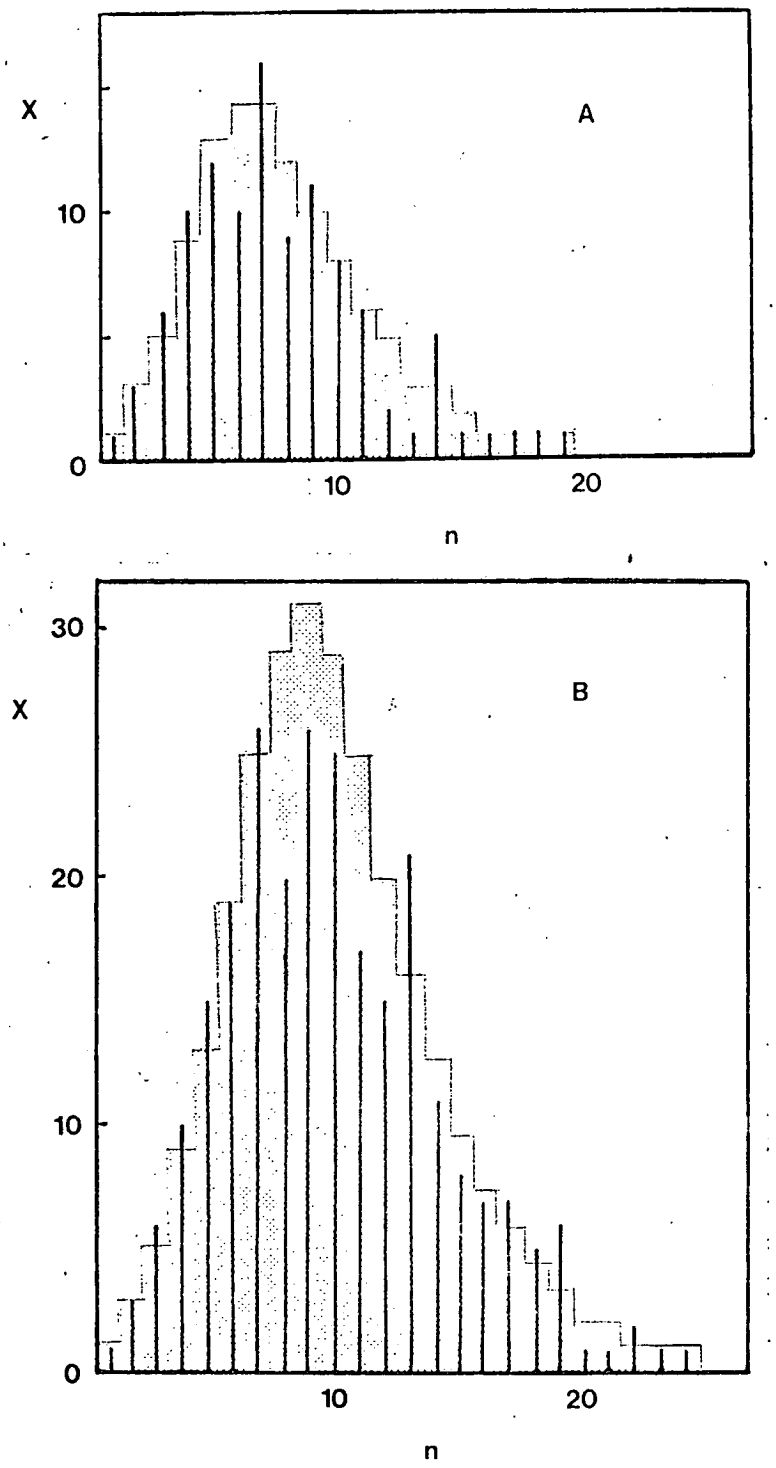


Fig.7.3. Compositional isomer vs. isostich plots. The hatched areas are histograms calculated from equation (6) and the bars are the results from one simulation. A. T=2000, B. T=10000.

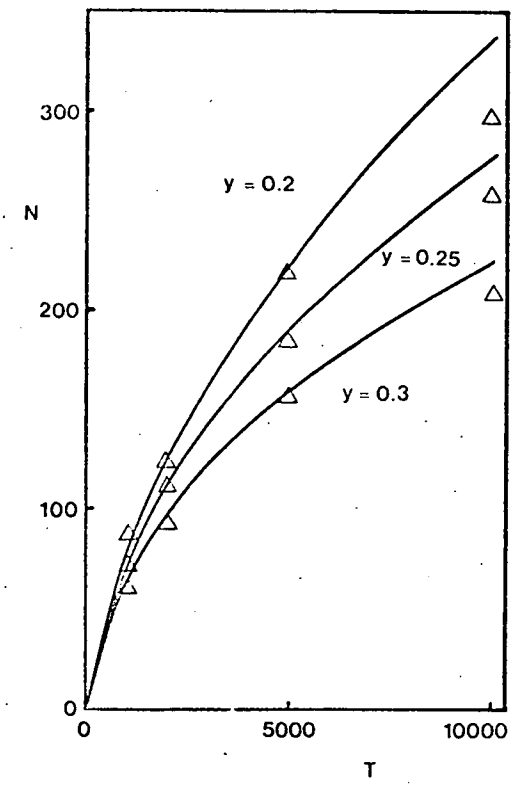


Fig.7.4. Total number of compositional isomers for three values of y. The curves are calculated from equation (7) and the points are the results of computer simulations. It can be seen that the relationship holds quite well up to a complexity of 5000.

the deviations from the expected plot do not become severe (χ^2 exceeded for 5% significance level) until $r \geqslant 0.5$.

Ø Function

The intensity data from the simulated digests is plotted in the form of histograms of spot number in classes of width 0.002 normalized intensity units. Fig.7.5 shows predictions and results for ideal sequences with complexities of 2000 and 5000. The predictions of equation (8) can be seen to be not bad, but they overestimate the number of spots in the lowest intensity class. The equation predicts that the effect of altering y will be mainly on the total spot number (the area under the histogram) rather than on the mode or the dispersion, and this was indeed found to be the case. Simulated intensity spectra from the Markov chain sequences showed severe deviations from the prediction outside the range $0.2 < r < 0.3$.

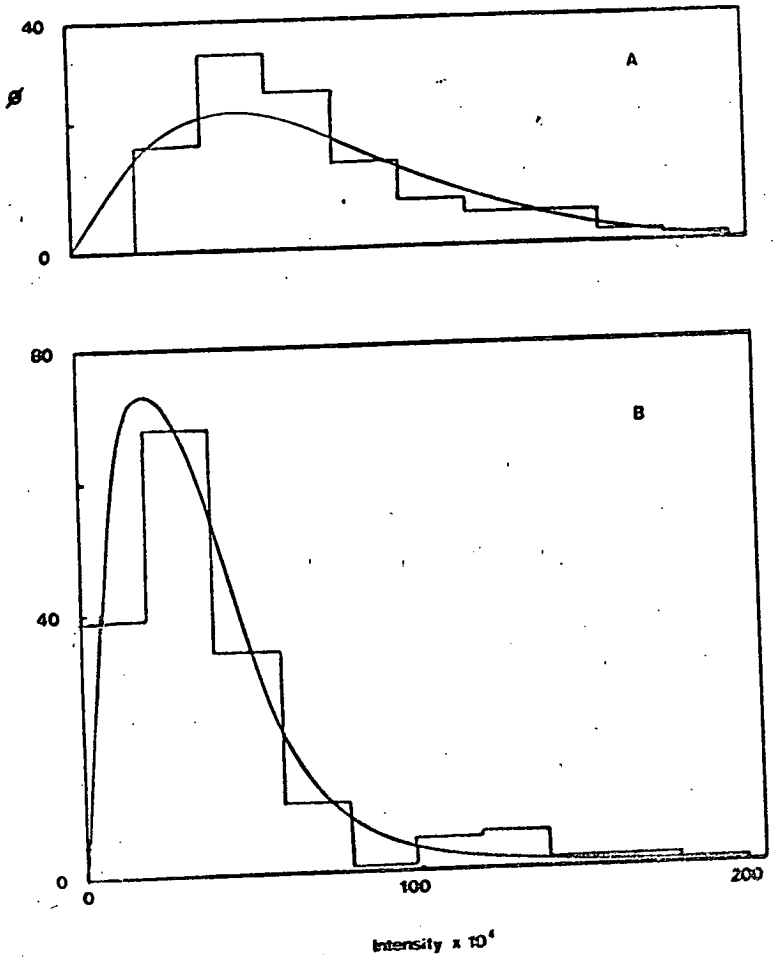


Fig.7.5. Intensity spectra. The curves are calculated from equation (8) and the histograms are from computer simulations. A. T=2000, B. T=5000.

The conclusion from this simulation study of single sequences is that the equations derived in the previous chapter are at least valid for predictive purposes within the ranges $1000 < T < 5000$, $0.2 < r < 0.3$ and $0.2 < y < 0.3$ and may be usable outside these limits under defined conditions.

Experiments on Xenopus rRNA

Looked at in one way, experiments are simply a slower and less clearcut way of doing computer simulations. However, once the equations have been verified by simulations, there arises an additional reason for carrying out some of the relevant measurements on naturally occurring RNAs, because any deviations from predicted behaviour may tell us something about the sequences concerned.

The first question concerns the randomness of the sequences of ribosomal RNA. Presumably the function of ribosomal RNA at least involves the formation of a precise three dimensional structure to hold the various ribosomal proteins in the correct positions. If this leads to significant non-randomness in the primary sequence, it should be detectable by means of the log(z/n) vs. n-l plot. The randomness referred to here is independence of successive positions in the sequence, ribosomal RNA is not strictly ideal because it does not have the four bases present in equal proportions.

Secondly, we can ask whether the molecules have fingerprints showing the expected number of spots. Too few spots might suggest internal duplications of sequence, and too many might suggest sequence heterogeneity caused by divergence between copies of the ribosomal genes. It will be recalled that although this factor was excluded as the cause of the heterogeneity of 40s RNA (chapter 3), a constant, low level of divergence is required by several of the theories of gene conservation (chapter 1) and is hard to rule out experimentally.

The predictions for the 18 and 28s RNAs of Xenopus were made using the following values for sequence length and base composition:

		18s RNA	28s RNA
		T = 2120	T = 4550
molar	A	0.239	0.194
proportions	C	0.243	0.282
of bases	U	0.233	0.171
	G	0.285	0.353

These are all taken from reference 14.

Isostich separations

Xenopus cells were labelled for 48 hours with $^{32}P_i$ and 18 and 28s RNAs were prepared as described in chapter 2. The RNA was digested with pancreatic RNAse (RNAse A) in 10 mM TrisCl pH 7.4, 1 mM EDTA for 1 hour at 37°, with an E:S ratio of 1:2. The samples were brought to 7 M urea and were chromatographed on columns of DEAE Sephadex. The procedure was the same as that described in chapter 4 for the fractionation of alkaline hydrolysis products, except that the column size was increased to 0.6x30 cms and the gradient was 0-0.5 M NaCl. 2 ml fractions were collected in polythene tubes and counted by Cerenkov radiation.

These columns gave quite good resolution of the isostichs, as is shown in Fig.7.6. The proportion of radioactivity in each peak is the quantity z(n) and so the values of log(z/n) could be readily calculated. These values for 18 and 28s RNA are shown plotted in Fig.7.7 over the theoretical lines calculated from equation (2) and the error envelopes calculated from equation (5). It can be seen that the agreement between

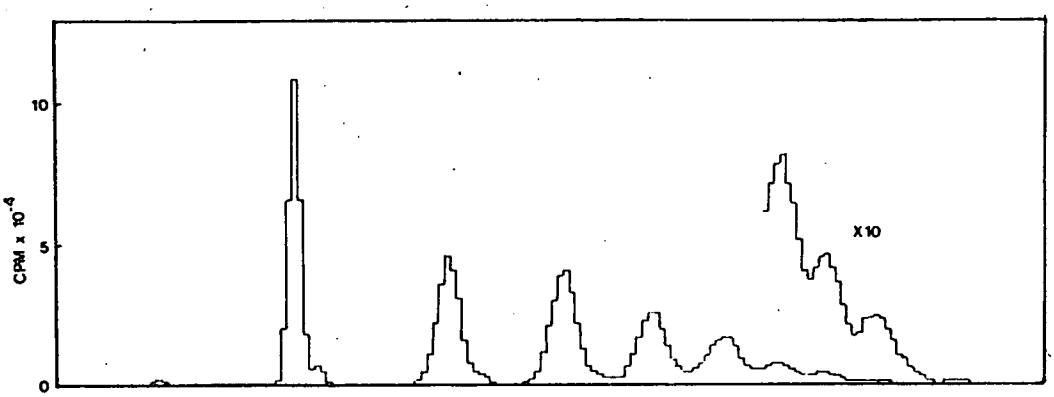


Fig.7.6. DEAE Sephadex column. Separation of RNAse A isostichs from Xenopus 28s RNA.

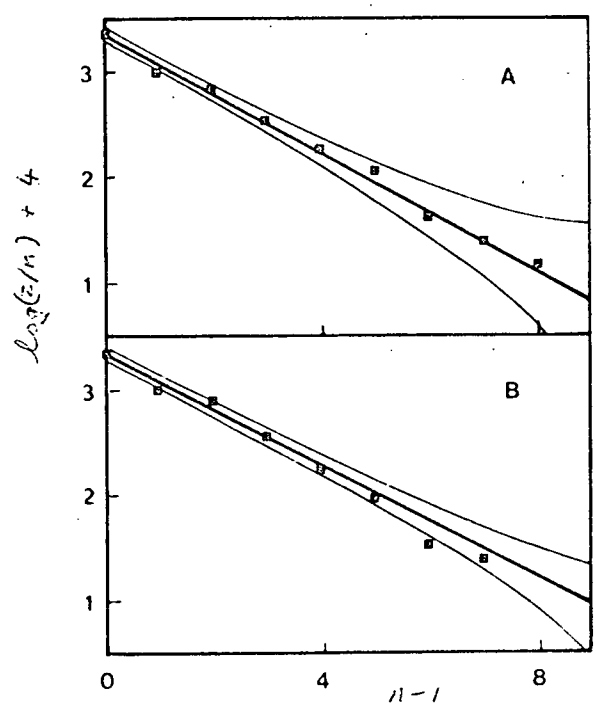


Fig.7.7. Log(z/n) vs. n-1 plots for Xenopus rRNAs. A. 28s, B. 18s. The lines are calculated from data in ref. 14 and the points are from the experimental isostich separations such as that of Fig.7.6.

theory and experiment is very good, and that there is therefore no reason to suppose that either of the RNA species has a significantly non-random sequence.

Fingerprints

These were carried out by the procedure in use in the laboratory of E.M.Southern. Samples of 18 and 28s RNA of high specific activity (6×10^5 cpm/μg) were desalted by passage through 1 ml columns of Sephadex G-75 and dissolved in 10 mM Tris Cl pH 7.4, 1 mM EDTA. They were digested with T_1 RNAse for 30' at 37° at an E:S ratio of 1:20. 2 μg of a digest was used for each fingerprint.

The first step was electrophoresis on cellulose acetate (Oxoid) for 3 hours at 5 kV. The buffer was 5% pyridine acetate pH 4.3 in 7 M urea. The sample was applied together with marker dyes and the strip was run immersed in paraffin to avoid overheating. After running, the strip was washed in ether and the contents transferred to large (16" x 14") sheets of PEI cellulose by ascending chromatography in water. These thin layers were not the commercial ones used previously, but were home made on plastic sheets of film backing by the following method: 30 grams of MN cellulose were mixed with 200 mls of 1% polyethylene imine + formic acid pH 6.5, the mixture was homogenized for 2' to remove lumps and was degassed. The sheets were made by pulling the backing sheet under a spreading trough with 0.4 mm aperture. They were allowed to dry and were left for a few weeks to mature. After the transfer, the layers were washed in water to remove urea and were chromatographed overnight in 2 M pyridine formate pH 1.8. They were dried at 80° and the spots were visualized by autoradiography.

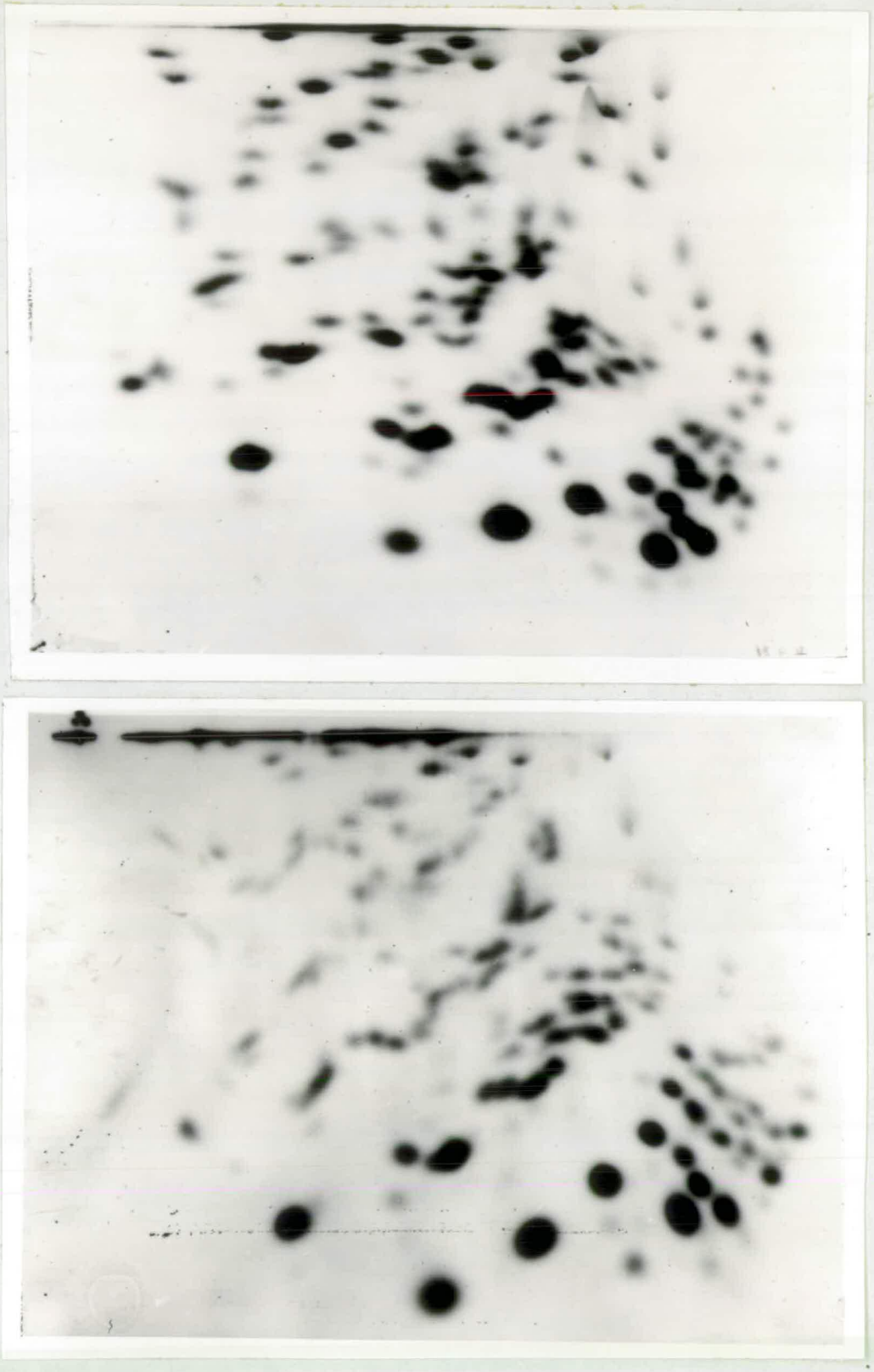


Fig.7.8. "Fingerprints" of Xenopus rRNAs carried out by the method described in the text. Top: 18s RNA, bottom: 28s RNA.

The thin layer modification of the original fingerprinting technique was used because of its superior resolution. It was hoped to separate all the compositional isomers from the Xenopus rRNAs, and in the best fingerprints this was virtually achieved (Fig.7.8). The spots were counted by comparing several fingerprints of the same RNA, because some resolve a particular region better than others. The final figures obtained were 122 for the 18s RNA and 130 for the 28s RNA.

These numbers were slightly higher than the theoretical predictions, so some additional computer simulations were carried out to produce confidence limits for the predictions. Ten random sequences were generated with the length and base composition of the two rRNAs and were fingerprinted by means of the AJSSAT program. The results are summarized below:

	Spot numbers	
	18s RNA	28s RNA
Calculated	106	125
Simulation	105 ± 9	118 ± 11
Experimental	122	130

At present the opinion of the author is that the slight excess of spots is a technical artifact and is not because the sequences have a higher complexity than predicted from their molecular weight. The main reason is probably the separation of some sequence isomers, since some such separations are plainly visible eg ACG and CAG.

Discussion

Some people find it hard to believe that an RNA sequence can be "random", or that a theory based on an ideal sequence can be at all valuable. This presumably springs from the reductionist view that a sequence which has a specific function eg coding for a protein, must also have a specificity at a lower level of organization, namely the primary sequence. The author believes that necessity at one level may be chance at another. Of course it is impossible to maintain that the whole genome is random because DNA can be fractionated into distinct classes of base composition or nearest neighbour frequency. But this does not deprive the ideal sequence of its uses, and in fact the next chapter consists of a theoretical study of one well known case of non-ideality, that of the repeated diverged sequence.

As far as the results described in this chapter are concerned, the simulation study clearly verifies the main predictions of the theory for ideal sequences and for non-ideal sequences within a defined range. The isostich separations for Xenopus rRNAs show experimentally that these species also follow predictions made from the assumption of a random sequence of their base composition.

It seems from the fingerprinting experiments described above that at present this technique is not up to accurate measurements of complexity because of the problems of non-specific digestion and the separation of sequence isomers. However, the theory does explain one striking feature of these results. It may have surprised some readers that the number of spots was so similar for the 18 and 28s RNAs even though the latter is more than twice as long as the former (4550

and 2120 nucleotides. But in this respect the predictions were quite correct. The explanation in words is that the lower G content of the 18s makes the oligonucleotides of a longer average length, so although there are far fewer copies, they are distributed over more compositional isomers. So, the similarity in spot number should not be taken to indicate, for example, that the 28s is a slightly diverged dimer of the 18s (135).

Chapter 8

FINGERPRINTS OF DIVERGED REPEATED SEQUENCES

Introduction

The properties of and the problems associated with the multiple gene arrays were discussed in chapter 1. Whether or not the multiple copies of such arrays as the ribosomal genes are diverged depends on which theory of their mechanism of evolution you believe in. But there is no doubt that there is divergence within the satellite DNAs (40) and immunoglobulin V genes(19) not to mention the kinetically defined reiterated DNA. From a theoretical point of view, such diverged repeated sequences represent a gross deviation from the ideal sequence defined above. So a special study has been made of the expected properties of such arrays and of the digests of RNA transcripts from them.

The model with reference to which the discussion will be conducted is the simplest possible one illustrated in Fig.8.1. The sequence family is a set of sequences derived from a single ancestral sequence by an M fold multiplication and an accumulation of random substitutions. A microsequence will be defined as a single member of the family. A microsite is a single position in a microsequence and a macrosite is a complete set of homologous microsites. The divergence, D, will be defined as the probability that a given macrosite in a set of two microsequences contains different residues. The divergence is a property of the family as a whole and not of individual microsequences.

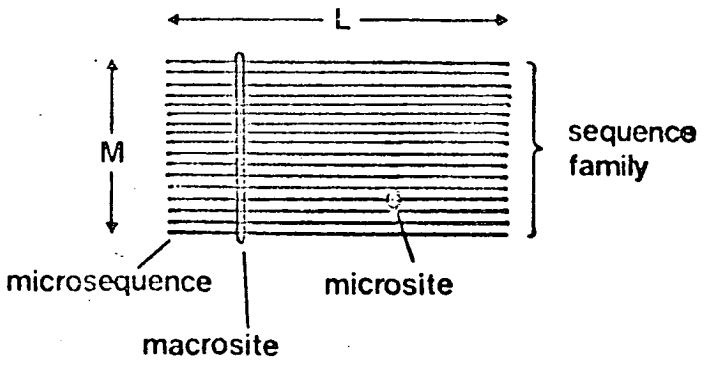


Fig.8.1. A simple sequence family.

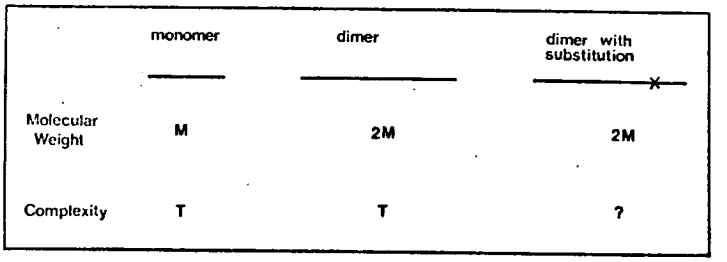


Fig.8.2. Illustration of the problem of defining "complexity" for a diverged repeated sequence.

In chapter 6 the complexity, T, was defined as the total unique sequence length. Such a definition is not satisfactory when we are dealing with diverged repeated sequences. This may be illustrated when we consider the three molecules in Fig.8.2. The complexity of the third molecule is clearly neither T nor 2T but something in between. In order to be able to assign a value, it is here defined by means of a rearrangement of equation (7):

$$T = \frac{1}{y^2(1-y)^{n-1}} \log \frac{\left(1 - \frac{1}{f(n)}\right)}{\left(1 - \frac{X(n)}{0.9 f(n)}\right)} \quad \ldots\ldots (9)$$

In practice, T will be measured from the number of spots on a fingerprint by using the curves of Fig.6.3.

Relationship Between Extent of Substitution and Divergence

In the computer simulations of the evolution of a sequence family (next section) we know how many random substitutions have been introduced overall but we do not know the divergence, which is what is measured by experimental techniques such as melting point depression of a denatured and reassociated sample of DNA. A relationship between the divergence, D, and the total number of substitutions, t, can readily be found if we define a new variable called "the virtual frequency of mutation", m. It is necessary to assume that all the twelve possible residue replacements occur with equal frequency (an approximation, see ref. 136) and that there are in addition null events (eg A→A) of the same frequency. Then for M sequences of length L:

$$m = \frac{4t}{3LM}$$

The distribution of mutated sites represents a classical occupancy problem under conditions in which a Poisson distribution is applicable (ie when mL/L^2 is small). For a single microsequence, the proportion of unmutated sites will therefore be e^{-m}. For two microsequences, the proportion of unmutated macrosites will be e^{-2m}. Any pair of microsequences should show the same degree of difference, and because of the definition of m both singly and doubly mutated macrosites have the same probability of an interstrand difference (viz 0.75). Therefore:

$$D = 0.75\,(1 - e^{-2m}) \quad \ldots\ldots \quad (10)$$

and this equation was used to calculate the values of the divergence in the simulations to be described below.

Simulation of Digests of Reiterated Sequences

The question to which we address ourselves in this section is: can we distinguish a unique from a repeated sequence by means of its fingerprint ? If so, can we calculate the constants which define the sequence family, namely L, M and D ?

This has been investigated by the method of computer simulation. The program used was similar to AJSSAT, which was used to produce the single sequence digests described in the previous chapter. But in this case, the data sequence was multiplied M times and random substitutions were introduced into the sequence family. At various prespecified times a digestion of the entire family was carried out.

What happens is illustrated in Fig.8.3. This shows a few of the compositional isomers from repeated fingerprints of a family with dimensions L=50, M=100. The horizontal axis represents the spot intensity.
'a' represents an increasing number of new spots which were brought into existence by the mutational events and build up to an equilibrium intensity.
'b' is a spot which remains fairly constant in intensity since it is close to equilibrium both for the starting and for the final complexity.
'c' and 'd' are ancestral sequence spots which are falling in intensity to the new equilibrium values.
'e' is a surprise spot which achieves an anomalously high intensity and which subsequently drops back again.
It may be seen that the behaviour of the spots in this simulation is fairly complicated, and these are only five out of a large number of compositional isomers in the various digests. It will in fact be more profitable to examine the problem in terms of the z, X and $\emptyset$

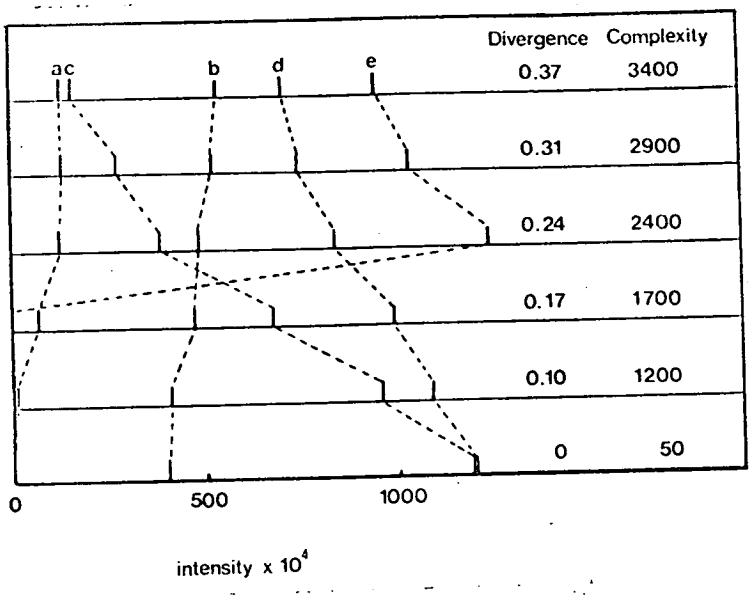


Fig.8.3. How the intensities of particular spots in a fingerprint of a sequence family change with divergence.

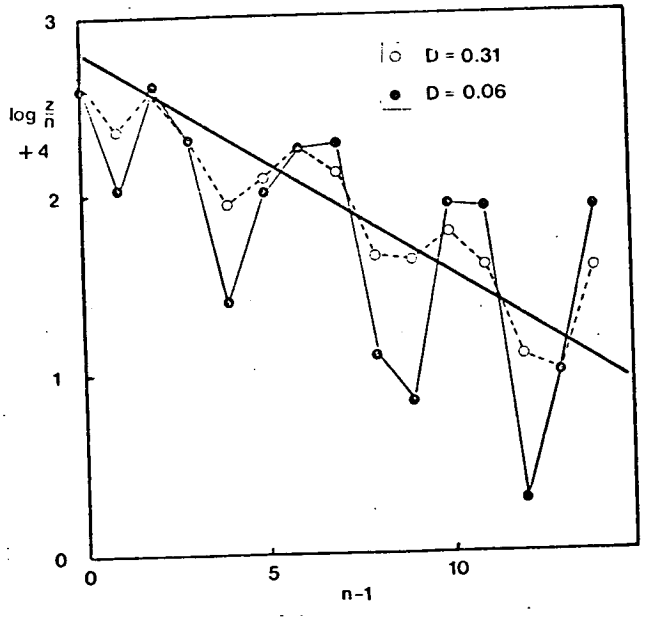


Fig.8.4. Linear plots for isostichs from a sequence family at two different degrees of divergence.

functions derived in chapter 6.

It was found that the log(z/n) plot was quite anomalous for such a digest. In Fig.8.4 are shown the points for a sequence family (100x100) at two different degrees of divergence. The scatter of points was very great and as the divergence increased it became progressively reduced to that appropriate to the complexity LM, in this case 10000. Unfortunatly the main features of this kind of plot are accidental ones which depend on the nature of the ancestral sequence, and so not much can be deduced from it except that the sample is non-ideal.

What about the spot number and the spot intensities of a fingerprint ? This will be clearer if the intensities are considered first. When the intensity spectra were plotted for different divergences it was found that the mode was always in a position appropriate to the complexity LM, as shown in Fig.8.5. The reason for this is that above a very small divergence the main sequence spots make little contribution to the spectrum because they are so few in number. So the position of the intensity maximum should make it possible to measure LM experimentally.

The total spot number, N, was found to rise with divergence from that appropriate to a complexity L to that appropriate to a complexity LM. From various simulations it was concluded that for low values of the divergence there is a linear relationship between T and D of the form:

$$T \simeq 2LMD$$

Therefore it should be possible to estimate the divergence of a family from its fingerprint as follows:

/over

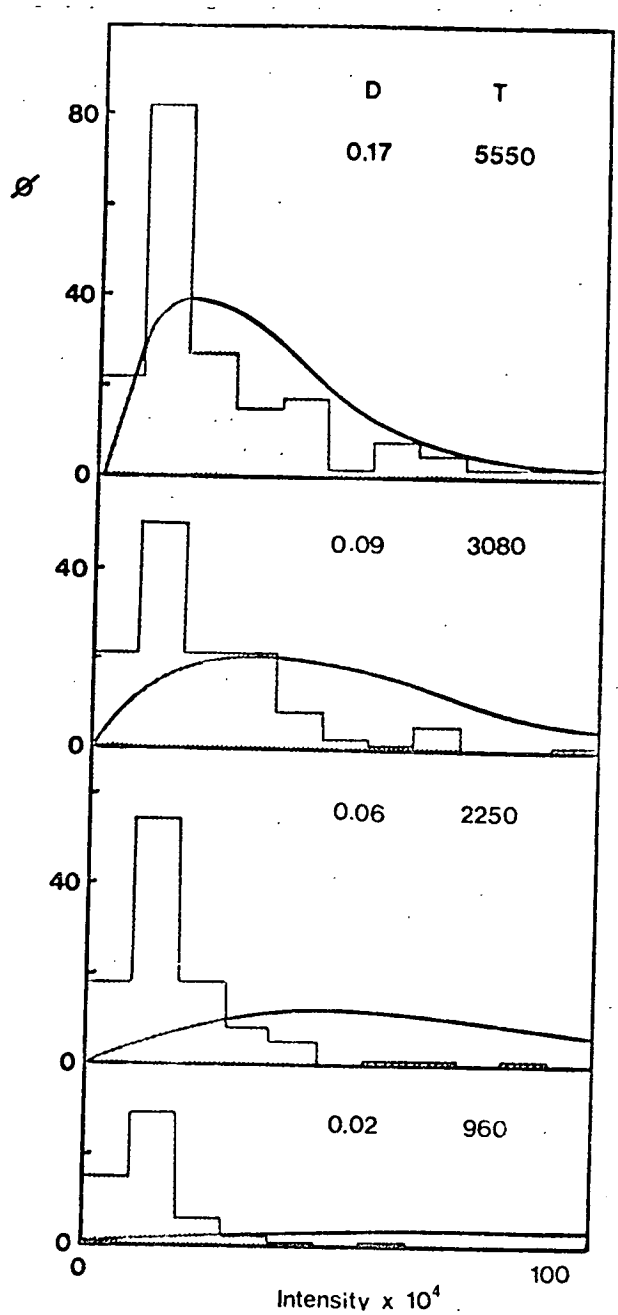


Fig.8.5. Intensity spectra for sequence family digests. The curves show the theoretical spectra appropriate to the complexities exhibited by the digests. The histograms show the actual spectra. It can be seen that the mode of the distribution is in a constant position which is that appropriate to the complexity 10000 (=L x M).

T is found from the total spot number by means of the curves of Fig.6.3. LM is found from the mode of the intensity spectrum; and then D is approximately equal to T/2LM.

Discussion

These results suggest that, in a favourable case, the analysis of a single fingerprint should enable measurements to be made of L, M and D. There are of course factors which may militate against such an analysis. For example, a family with a very large number of copies will have minor spots of such low relative intensities that they will probably pass unnoticed and the sequence be regarded as unique. There may also be families that have come into existence by a more complex process than multiplication and divergence, and in such a case this treatment would not be appropriate.

When we examine the fingerprints of Xenopus rRNAs in the light of this treatment we can draw certain conclusions. The fingerprints of 5s RNA from the kidney and ovary are different (31,32) the formaer having 22 spots and the latter 41. Various sites in the ovary sequence were found to be heterogeneous, eg to have 50% of each of two different nucleotide residues. This behaviour is not characteristic of the sort of sequence family illustrated in Fig.8.1, because there are about 20,000 copies of 5s DNA per haploid set (18) and so vastly more spots would be expected before any single site became appreciably substituted. It is much more likely that this heterogeneity, together with the melting point depression in the reassociated DNA (18) is due to differences between gene clusters. Only one cluster would be transcribed in the kidney cells and

several would be transcribed in the ovary, which is very active in 5s RNA synthesis.

The fingerprints of the 18 and 28s RNAs, which were carried out by the author and described in the previous chapter, gave a small excess of spot number over the expected values. What are the chances that this excess is significant and is caused by a low level of divergence among the ribosomal genes ? In the author's opinion they are slim. Although he believes that there is a small amount of divergence in the ribosomal genes, the same argument holds as for the 5s RNA. A small amount of divergence would give very faint spots of about 1/600th the average intensity, while a large amount of divergence would give far more spots than were observed and would also be detectable by other methods. Since there is only one ribosomal locus per chromosome set in Xenopus there is no a priori reason to expect divergence between clusters which might give rise to a few extra spots at high intensity.

Chapter 9

POSSIBLE EXTENSIONS TO THIS WORK

The Cycloheximide Effect

In this thesis the effect is described but not investigated. It is not known whether the accumulation of 40s RNA is caused by the inhibition of protein synthesis or by some specific effect of the drug. The latter possibility would be excluded if it occurred with other protein synthesis inhibitors, and there is some evidence that this is the case for puromycin (David Hutchens, Zoology Honours Project). It would of course be interesting to identify the proteins which are required, the most obvious possibility being that they are ribosomal proteins.

Secondary Structure of RNA

Probably the simplest method of proving the conclusion of chapter 3 would be to compare partial ribonuclease digests of material from the right and left hand sides of a gel electrophoresis peak. If there are several metastable arrangements of the base paired loops in the sequence then there might be some qualitative differences in the partial digests. However, a negative result from such experiments would have little significance.

The best proof would be to demonstrate by computation that the 40s RNA will form a variety of metastable conformations. But this will have to await the publication of the complete sequence, for at present we have only the 5' end.

"Read through" of Ribosomal Genes

It was implied in the discussion following chapter 5 that the author favours a mechanism for ribosomal RNA synthesis in which there is no reinitiation for each repeating unit, and the polymerase moves through the spacer by some process akin to transcription. One confirmation of this idea would be to show a conversion of nucleoside triphosphates to 5' monophosphates by polymerase I + ribosomal DNA, or by *in vitro* nucleoli. This process should be sensitive to actinomycin D.

The Trimethylated Sequence

The importance of this is that it is an easily recognised marker in the 28s sequence. We have in fact already used this property to check the fidelity of methylation by isolated rat liver nucleoli by showing that the tetranucleotide is labelled by (^{3}H-methyl) S-adenosyl methionine.

Before the publication of the electron microscopic map of the 40s RNA, it was thought that the order of the 18 and 28s sequences could be determined by exonuclease digestion of the precursor, using the alkali stable tetranucleotide as a marker. Some trial experiments were performed using polynucleotide phosphorylase, but these were abandoned because of the problem of endonucleolytic degradation.

The Spot Number/Complexity Relationship

This is likely to be the most useful of the equations derived in chapter 6 because it provides a method of measuring the complexity of mixed RNAs by means of a single fingerprint. It is a different sort of measurement from that provided by hybridization.

An RNA excess hybridization shows what proportion of the DNA is complementary to the RNA, and a DNA excess hybridization shows which kinetic fraction of the DNA is complementary to the RNA, but neither gives the complexity of the RNA. The fingerprint method is independent of the relative amounts of the different species so long as all the digestion products are detectable and resolvable.

Three possible applications of interest are:
i) Measurement of the complexity of total HnRNA or total mRNA under various circumstances.
ii) Measurement of the complexity of the host sequences associated with SV 40 mRNA, which is under study in this laboratory, to find whether the virus integrates at a few sites or at random.
iii) Checking the fidelity of transcription by E.coli RNA polymerase *in vitro*.

Sequence Families

The theory of sequence families presented in chapter 8 will have to await some suitable data for its application. This data would consist of high quality fingerprints of *in vitro* transcripts of DNA. An attempt was made to analyse such data obtained by A.Biro for mouse satellite DNA. This was unsuccessful, but this is not surprising because the sequence evidence strongly suggests a multistage multiplication and divergence which means that the model of chapter 8 is not applicable.

APPENDIX I

An Improved Method for the Preparation of Polynucleotide Kinase

Polynucleotide kinase is an enzyme produced by T-even phages in the early part of the infective cycle. The standard preparative method was described by Richardson (137), and the yields have more recently been increased by using a phage mutant blocked in DNA replication (138). Such mutants form about three times the normal amount of enzyme and because they cannot complete maturation, the timing of cell disruption is not critical.

The enzyme was prepared twice for the work described in this thesis. The protocol below describes the second, and simpler, procedure, which differs considerably from that of Richardson.

1. Growth of Phage

E.coli CR 63 (amber suppressing) was grown to a turbidity of 0.5 (about 0.5×10^9 cells/ml) on L broth in a 10 l. fermenter. The phage lysate (T4am82) was added at a multiplicity of 10:1 at which point the mixture frothed considerably. After 1½ hours a few drops of chloroform were added to the lysate which was stored in the fridge.

2. Growth of Infected Cells

E.coli B/r/T1 (non-suppressing) was grown on L broth in a 50 l. fermenter to a density of $0.8\text{-}1 \times 10^9$. The lysate prepared in the first stage was added at a multiplicity of 4:1 accompanied by 10 μg/ml of

tryptophan. The infection was allowed to continue for 90' after which the temperature was reduced to 10° and the cells spun down in a continuous centrifuge. They were stored frozen at -20°.

3. Cell Disruption

50 g. of frozen cells were ground for 10' with 100 g. of alumina (Alcoa). The sludge was mixed with 250 mls of 50 mM Tris Cl pH 8, 0.2 M NaCl, 10 mM mercaptoethanol, 1 mM EDTA and 5% glycerol, and centrifuged at 10K for 20'. The pellets were reground for 5' and then recentrifuged. The pooled supernatants were centrifuged again at 28.5K for 2¼ hours in the Spinco 30 rotor and the pellets were discarded. At this stage it was estimated that the solution contained about 80% of the soluble protein from the cells.

4. 1st Ammonium Sulphate Precipitation

The solution was brought to 50% saturation by dropwise addition of saturated ammonium sulphate pH 7.4 at 0°C. The suspension was stirred for 30' and the precipitate spun down at 10K for 15'. It was dissolved in 100 mls of 10 mM Tris Cl pH 7.4, 10 mM mercaptoethanol, 5% glycerol. This was diluted until the conductivity had fallen to a value equal to that of 0.3 M NaCl in the same buffer.

5. Removal of DNA

A 250 ml column of DE-23 cellulose was equilibrated in 10 mM Tris Cl pH 7.4, 10 mM mercaptoethanol, 5% glycerol, 0.3 M NaCl. The sample was run onto the column and was followed by the same buffer. At this salt concentration, the column should retain nucleic acid fragments but allow the enzyme to pass through.

The A_{260}/A_{280} was measured for the fractions, and those with a ratio of less than 0.8 were pooled for further purification.

6. 2nd Ammonium Sulphate Precipitation

The solution was brought to 50% saturation of ammonium sulphate as before and the precipitate was spun down at 10K for 15'. It was dissolved in 24 mls of 50 mM sodium phosphate pH 7.4, 10 mM mercaptoethanol and dialysed against three changes of one litre of the same buffer to remove the salt. The precipitate which formed during dialysis was spun down at 10K for 15' and was discarded.

7. Phosphocellulose Column

A 50 ml column of phosphocellulose was equilibrated in 50 mM sodium phosphate pH 7.4, 10 mM mercaptoethanol. The sample was run on and washed with 40 mls of buffer. Then two "steps" were applied, the first being 80 mls of buffer + 0.1 M NaCl and the second being 120 mls of buffer + 0.25 M NaCl. All the fractions were assayed and the main peak of activity was found to be associated with the second step (Fig.A.1). These fractions were pooled and were dialysed overnight against 10 mM sodium phosphate pH 7.4, 10 mM mercaptoethanol.

8. DEAE Cellulose Column

The 20 ml column was equilibrated in 10 mM sodium phosphate pH 7.4, 10 mM mercaptoethanol. The sample was run on and washed with 40 mls of buffer, then a step was applied of 60 mls 50 mM sodium phosphate pH 7.4, 10 mM mercaptoethanol. The activity eluted with this step (Fig.A.2).

At this stage the preparation was ready for use.

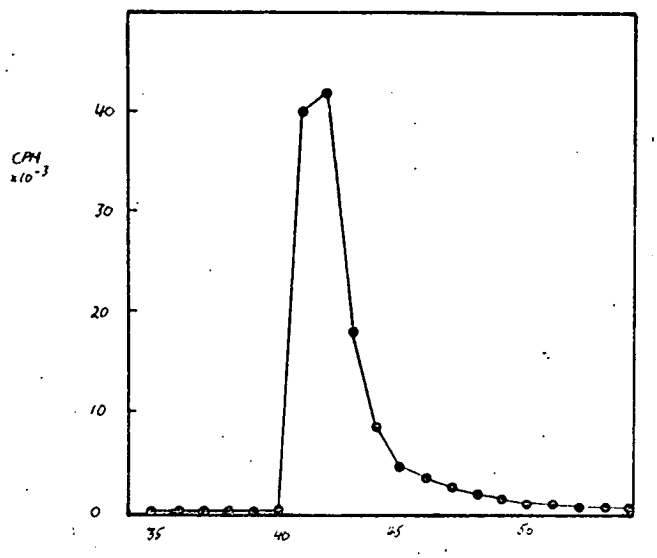


Fig.A.1. Phosphocellulose column chromatography of polynucleotide kinase.

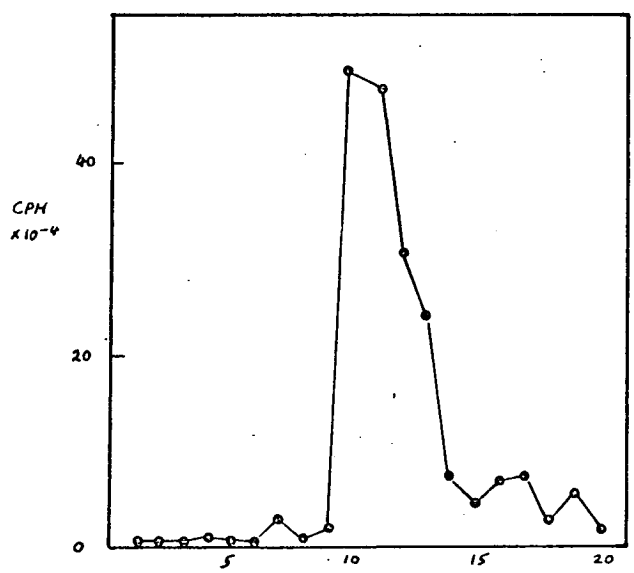


Fig.A.2. DEAE Cellulose column chromatography of polynucleotide kinase.

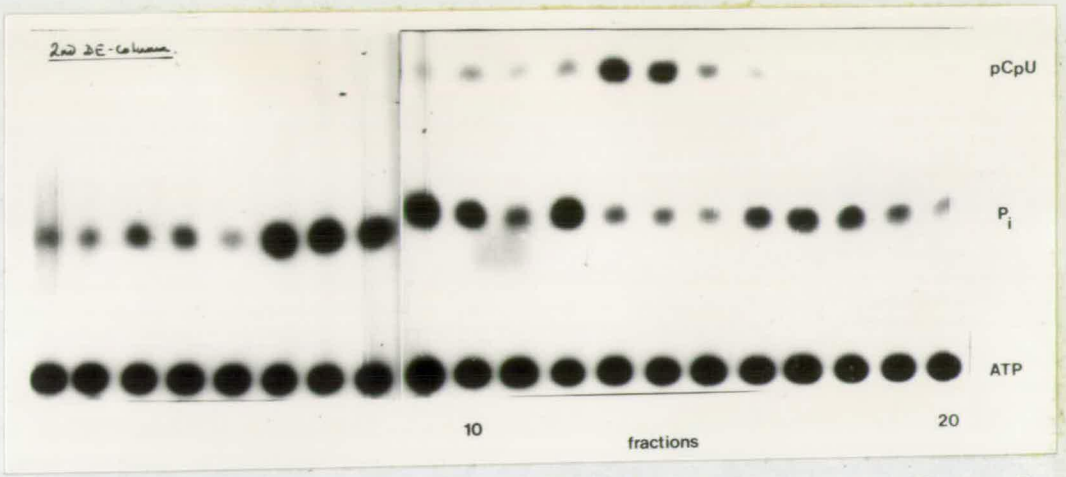


Fig.A.3. Assays of polynucleotide kinase performed by the method described in the text. This shows the assays for a number of fractions from a DEAE column.

Assays

The Folin method was used for protein (139). The enzyme assay was somewhat modified from a method used by K.Murray. The reaction mixture consisted of:

Tris Cl pH 7.4	20 mM
$MgCl_2$	5 mM
mercaptoethanol	20 mM
ATP	1 mM
γ^{32}P-ATP	trace
CpU	1.1 mM

5 µl. of this were mixed with 1 or 2 µl of sample and incubated under a plastic cap for 30' at 37°. Then it was transferred to a PEI cellulose layer and chromatographed for 30' in 1 M pyridine formate pH 3.5. The layer was dried and exposed to X-ray film for 1-2 hours. Thus visualized, the ATP and pCpU spots were cut out and counted.

Since the CpU is in excess over the ATP in the

reaction mixture, the extent of reaction is measured by ATP consumption. A 20% conversion of ATP to pCpU represents one unit of enzyme activity because 1 unit = 1 nmole transferred in 30' at 37°. Fig.A.3 shows a peak of enzyme activity demonstrated by this assay.

Purification Table

Stage	Total Protein (mg)	Total Activity (u.)	Spec. Act.	Fold Purif.
Crude extract	5040	7.2×10^5	1.43×10^2	1
1st. AS ppt.	540	1.3×10^5	2.4×10^2	1.7
DEAE I	460	8.4×10^4	1.8×10^2	1.3
2nd. AS ppt.	178	1.1×10^5	6.2×10^2	4.2
Phosphocell.	33	7.5×10^4	2.3×10^3	16
DEAE II	5.4	5.4×10^4	1×10^4	70.

Storage

The first batch of polynucleotide kinase was stored at 4° in the cold room. Although this is the recommended method, the enzyme rapidly lost its activity. So a variety of storage methods were examined using the second batch. It was found that freezing and thawing, however fast, caused almost total loss of activity. Inclusion of 10% DMSO or glycerol before freezing gave good protection. The DMSO is probably better but it does give rise to a foul smell in the solution, probably by oxidation of mercaptoethanol. 80-100% of the enzyme activity remained after one cycle of freezing/thawing in DMSO.

APPENDIX II

Computer Programs

A number of computer programs were written to carry out the work described in chapters 6, 7 and 8. All were written in FORTRAN and were implemented on the Edinburgh Multi Access System. The programs and data are held on the magnetic disc and may be edited or run from a remote console. The actual listings of the programs can be provided on request, but this appendix should give an idea of what each one does.

Programs for Mathematical Computation

1. AJSSLP: This program is supplied with the values of y and T for a sequence and calculates the constants for the linear plot and the confidence limit envelopes:

$$\text{slope} = \log(1-y)$$

$$\text{intercept} = 2\log y + 4$$

$$\text{limits} = \pm 1.96 \log_{10} e \left[\frac{n(1-z)}{Tz}\right]^{1/2}$$

Data:

y	NN	II
0.250	20	3

T
1000
2000
10000

y = proportion of break points.
NN = number of isostichs used in calculation.
II = number of values of T used for envelope computation.
T = complexity.

Sample output:

```
SLACK PLOT FOR Y = 0.250
Y-AXIS + 4 UNITS
Y-INTERCEPT = 2.796      SLOPE = -.125

ENVELOPES:
T =  2000
N-1    95% LIMITS
 0    2.72    2.87
 1    2.59    2.75
 2    2.45    2.64
 3    2.31    2.53
 4    2.17    2.42
 5    2.02    2.32
 6    1.87    2.22
 7    1.72    2.12
 8    1.56    2.03
 9    1.40    1.94
10    1.23    1.86
11    1.06    1.79
12    0.87    1.72
13    0.68    1.66
14    0.48    1.61
```

2. AJSNUC: This calculates values of X(n) and sums them over the isostichs to provide an estimate of the number of spots on a fingerprint:

$$X(n) = 0.9\, f(n) \left[1 - \left(1 - \frac{1}{f(n)} \right)^{R} \right]$$

$$N(T) = \sum_{n=1}^{n=n_1} X(n)$$

The data is inserted into DATA statements within the program and consists of values for y, T, n_1 and whether the digestion is by RNAse A or RNAse T_1.

Sample output:

THEORETICAL CURVES FOR RNASE T1 DIGESTION

PROPORTION OF BREAK POINTS = 0.250

COMPLEXITY = 5000 NUCLEOTIDES

ISOTICH	NO.COMPS.
1	1.0
2	3.0
3	6.0
4	10.
5	15.
6	20.
7	24.
8	25.
9	23.
10	19.
11	16.
12	12.
13	9.4
14	7.2
15	5.5
16	4.1
17	3.1
18	2.3
19	1.8
20	1.3
21	0.99
22	0.74
23	0.56
24	0.42
25	0.31
26	0.24
27	0.18
28	0.13
29	0.99E-01
30	0.75E-01

TOTAL NUMBER OF SPOTS = 210

3. AJSINS: This calculates intensity spectra. The calculation proceeds independently for each isostich and the expression used depends on the value of R. If $R < 50$, the binomial formula is used:

/over.

$$\phi(n,I) = \frac{X.R!}{\left(R-\frac{TI}{n}\right)!\left(\frac{TI}{n}\right)!}\left(\frac{1}{f}\right)^{\left(\frac{TI}{n}\right)}\left(1-\frac{1}{f}\right)^{R-\left(\frac{TI}{n}\right)}$$

and when $R > TI/n$, $\phi = 0$. If $R \geq 50$, the normal approximation to the binomial is used:

$$\phi(n,I) \doteq \frac{0.395X}{\left[\frac{R}{f}\left(1-\frac{1}{f}\right)\right]^{1/2}} e^{-\left[\left(\frac{TI}{n}-\frac{R}{f}\right)^2 / \left(\frac{2R}{f}\right)\left(1-\frac{1}{f}\right)\right]}$$

All isostichs are summed to give the overall intensity spectrum.

Data:

```
      y
   0.250
   2000      20      1
    T        NN    AGAIN
```

y = proportion of break points.

T = complexity

NN = number of isostichs considered.

AGAIN = whether or not there is a further data set.

Sample output:

```
INTENSITY SPECTRUM FOR RNA FINGERPRINT T =  5550  Y = 0.250

FREQUENCY SUMMATION
     30.3        39.4        36.5        28.5        19.9
     13.0        8.19        5.23        3.50        2.49
     1.37        1.45        1.14       0.884       0.687
    0.554       0.485       0.458       0.446       0.421
MEAN = 0.0032  VARIANCE = 0.0651
```

Programs Used for Simulation

4. AJSSEQ: This program generates random sequences of the four letters A, U, C, G. The proportions of each are specified in the data and the letter for each site is selected by means of uniform random

numbers from a power residue random number generator.

Data:

2000	0.250 0.450 0.800	34534729	1
T	cumulative base proportions for A, U, C.	random starter	AGAIN

The output was written directly onto a disc file for use by one of the digestion programs.

5. AJSMAR: This generates random sequences which have base proportions all equal to 25% but which are Markov chains defined by a single association parameter r. For each site a random number is selected and the transition matrix is entered in the column of the previous base. The new base is chosen by comparing the random number with the four values in the column. All the diagonal elements are equal to r and all the non-diagonal elements to (1-r)/3.

Data:

2000	0.300	465785432	1
T	r	random starter	AGAIN

Again, the output is written directly onto a disc file.

6. AJSSAT: This takes a sequence supplied as data and produces an RNAse T_1 digest. The output consists of a table of the compositional isomers yielded by the sequence, in order of length, together with their intensities calculated with the formula:

$$\text{Intensity} = \frac{\text{length x incidence}}{\sum(\text{length x incidence})}$$

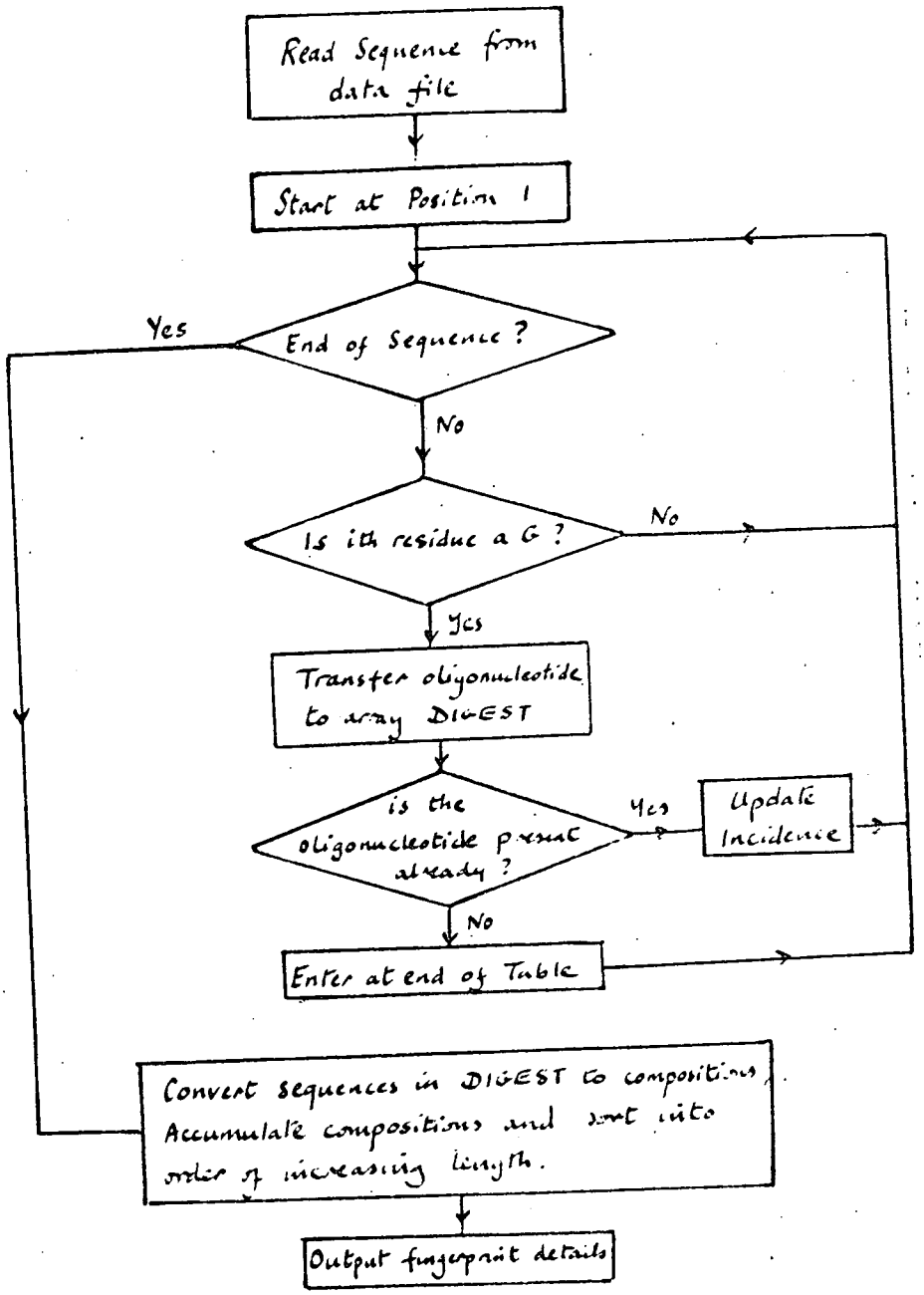


Fig.A.4. Flow diagram of the computer program AJSSAT, used to work out "fingerprints" from sequences supplied as data.

The principle of this program is very simple, but it does require some quite complicated sorting routines. The flow diagram is shown in Fig.A.4.

7. AJSUGH: This program is the same except that it operates on a sequence family. A short data sequence is duplicated M times and then random substitutions are introduced into the array. At certain prespecified extents of mutation, the whole array is subjected to the digestion procedure and a table of the compositional isomers is printed out.

Sample output:

```
SIMULATION OF EVOLUTIONARY DRIFT OF MULTIPLE GENES;
EFFECT ON T1 DIGESTION OF RNA
DATA: L =  48  N =  10
STARTING SEQUENCE IS:
GGGAUUGGGAUUGGGAUUGGGAUUGGGAUUGGGAUUGGGAUUGGGAUU
```

```
EXTENT OF MUTATION = 0.0800000
NUMBER OF SEQUENCES IN DIGEST =  21
NUMBER OF COMPOSITIONS IN DIGEST =  18
 FINGERPRINT DETAILS:
  U    A    C    G     LENGTH     INTENSITY
  0    0    0    1        1        0.3146
  0    1    0    1        2        0.0042
  2    0    0    0        2        0.0042
  1    0    0    1        2        0.0042
  1    1    0    1        3        0.0042
  2    1    0    0        3        0.0313
  2    0    1    1        4        0.0083
  2    1    0    1        4        0.4250
  2    2    0    0        4        0.0167
  1    2    0    1        4        0.0333
  3    1    0    0        4        0.0083
  2    1    1    0        4        0.0083
  1    1    1    1        4        0.0083
  2    1    1    1        5        0.0313
  3    1    0    1        5        0.0104
  2    2    0    1        5        0.0625
  1    1    2    1        5        0.0104
  1    3    1    1        6        0.0125
```

APPENDIX III

Materials

Radiochemicals

DL(^{3}H)leucine	15000 mCi/mmol
5-(^{3}H)-uridine	15000 mCi/mmol
(methyl-^{3}H)-methionine	5000 mCi/mmol
^{32}Phosphate	90 Ci/mg
(γ-^{32}P)-ATP	10000 mCi/mmol

All from the Radiochemical Centre, Amersham.

Enzymes

Alkaline phosphatase 30 units/mg
from Whatman Biochemicals Ltd., Maidstone, Kent.

Ribonuclease T_1 Grade III
Ribonuclease A Type 1 A
Deoxyribonuclease I Electrophoretically purified
Venom phosphodiesterase Type II
from Sigma Chemical Co. Ltd., Kingston-upon-Thames.

Separation Media

DEAE Cellulose, Whatman DE-23 & DE-52
from W.R.Balston Ltd., Maidstone, Kent.

DEAE Sephadex
from Pharmacia (GB) Ltd., Paramount House,
London W5 5SS.

PEI-Cellulose thin layers
from Camlab, Cambridge.

Oligo-dT Cellulose
from Collaborative Research Inc., Waltham,
Mass. 02154, USA.

Cell Growth Media

All from Biocult Laboratories, 3 Washington Road,
Paisley.

REFERENCES

Abbreviations:

PNAS	Proc.Nat.Acad.Sci. USA
JMB	J.Mol.Biol.
ECR	Exp.Cell Research
NNB	Nature New Biology
BJ	Biochemical Journal
JCB	J.Cell Biol.
JBC	J.Biol.Chem.
EJB	Eur.J.Biochem.
BBRC	Biochem.Biophys.Res.Comm.
ARB	Ann.Rev.Biochem.
ME	Methods in Enzymology
BBA	Biochem.Biophys.Acta
PNAR	Prog.Nuc.Acid Res. & Mol.Biol.

1. Elsdale T.R., Fischberg M. and Smith S. ECR 18 642-643 (1958).
2. Brown D.D. and Gurdon J.B. PNAS 51 139-146 (1964)
3. Yankofsky S.A. and Spiegelman S. PNAS 49 538-544 (1963).
4. Chipcase M.H.I. and Birnsteil M.L. PNAS 50 1011-1016 (1963)
5. Birnsteil M.L., Wallace H., Sirlin J.L. and Fischberg M. National Cancer Institute Monograph 23 431-447 (1966).
6. Clarkson S.G., Birnsteil M.L. and Serra V. JMB 79 391-410 (1973).
7. Brown D.D. and Weber C.S. JMB 34 681-697 (1968).
8. Birnsteil M.L., Speirs J., Purdom I., Jones K. and Loening U.E. Nature 219 454-463 (1968).
9. Perry R.P. PNAS 48 2179-2186 (1962).
10. Scherrer K., Latham H. and Darnell J.E. BBRC 7 486-490 (1962).

11. Scherrer K., Latham H. and Darnell J.E. PNAS 49 240-248 (1963).
12. Jeanteur Ph. and Attardi G. JMB 45 305-324 (1969).
13. Landesmann R. and Gross P.R. Dev. Biol. 19 244-260 (1969).
14. Loening U.E., Jones K. and Birnsteil M.L. JMB 45 353-366 (1969).
15. Miller O.L. and Beatty B.R. Science 164 955-957 (1969).
16. Wensink P.C. and Brown D.D. JMB 60 235-247 (1971).
17. Kedes L.H. and Birnsteil M.L. NNB 230 165-169 (1971).
18. Brown D.D., Wensink P.C. and Jordan E. PNAS 68 3175-3179 (1971).
19. Gally J.A. and Edelman G.M. Ann. Rev. Genet. 6 1-46 (1972).
20. Britten R.J. and Kohne D.E. "Handbook of Molecular Cytology", ed. Lima-de-Faria A. (N.Holland, Amsterdam) 21-51 (1969).
21. Church R.B. and McCarthy B.J. Biochemical Genetics 2 55-73 (1968).
22. Birnsteil M.L., Chipcase M. and Speirs J. PNAR 11 351-389 (1971).
23. Dawid I.B., Brown D.D. and Reeder R.H. JMB 51 341-360 (1970).
24. Brown D.D., Wensink P.C and Jordan E. JMB 63 57-73 (1972).
25. Brown D.D. and Sugimoto K. JMB 78 397-415 (1973).
26. Weinberg E.H., Birnsteil M.L., Purdom I.F. and Williamson R. Nature 240 225-228 (1972).
27. Birnsteil M.L., Weinberg E.H. and Pardue M.L. in "Molecular cytogenetics" ed. Hamkalo B.A. and Papaconstantinopolu J. (Plenum Press, N.Y.) 75-93 (1973).
28. Greenaway P.J. and Murray K. NNB 229 233-238 (1971).
29. Birnsteil M.L., Sells B.H. and Purdom I.F. JMB 63 21-39 (1972).

30. Clarkson S.G., Birnsteil M.L. and Purdom I.F. JMB 79 411-429 (1973).
31. Ford P.J. and Southern E.M. NNB 241 7-13 (1973).
32. Ford P.J. Biochem.Soc.Symp. 37 69-82 (1973).
33. Pardue M.L., Brown D.D. and Birnsteil M.L. Chromosoma 42 191-203 (1973).
34. Muramatsu M. and Busch H. Cancer Research 24 1028-1034 (1964).
35. Hashimoto S. and Muramatsu M. EJB 33 446-458 (1973).
36. Moore R.L. and McCarthy B.J. Biochemical Genetics 2 75-86 (1968).
37. Melli M., Whitfield C., Rai S.K., Richardson M. and Bishop J.O. NNB 231 8-12 (1971).
38. Ferguson-Smith M.A. Cytogenetics 3 124-134 (1964).
39. Birnsteil M.L. and Grunstein M. FEBS Symposium 23 349-365 (1972).
40. Walker P.M.B. Prog.Biophys.Mol.Biol. 23 145-190 (1971).
41. Callan H. J.Cell Sci. 2 1-7 (1967).
42. Smith G. Cold Spring Harbor Symp.Quant.Biol. 38 (1973). 507-513
43. Ritossa F.M. PNAS 60 509-516 (1968).
44. Bishop J.O. and Rosbash M. NNB 241 204-207 (1973).
45. Suzuki Y., Gage L.P. and Brown D.D. JMB 70 637-649 (1972).
46. Sullivan D. et. al. JBC 248 7530-7539 (1973).
47. Perry R.P. PNAR 6 219-257 (1966).
48. Georgiev G.P. PNAR 6 259-351 (1966).
49. Attardi G. and Amaldi F. ARB 39 183-226 (1970).
50. Loening U.E. Symp.Soc.Gen.Microbiol. 20 77-106 (1970).
51. Maden B.E.H. Prog.Biophys.Mol.Biol. 22 129-177 (1971).
52. Burdon R.H. PNAR 11 33-79 (1971).
53. Perry R.P. ECR 29 400-406 (1963).
54. Penman S., Smith I. and Holtzman E. Science 154 786-789 (1966).

55. Warner J.R., Soeiro R., Birnboim H.C., Girard M. and Darnell J.E. JMB 19 349-361 (1966).
56. Roeder R.G. and Rutter W.J. Nature 224 234-237 (1969).
57. Greenberg H. and Penman S. JMB 21 527-535 (1966).
58. Wagner E.S., Penman S. and Ingram V.I. JMB 29 371-387 (1967).
59. Warner J.R. and Soeiro R. PNAS 58 1984-1990 (1967).
60. Willems M, Penman M. and Penman S. JCB 41 177-187 (1969).
61. Loening U.E. BJ 102 251-257 (1967).
62. Bishop D.H.L., Claybrook J.R. and Spiegelman S. JMB 26 373-387 (1967).
63. Weinberg R.A., Loening U.E., Willems M. and Penman S. PNAS 58 1088-1095 (1967).
64. Weinberg R.A. and Penman S. JMB 47 169-178 (1970).
65. Wellauer P.K. and Dawid I.B. PNAS 70 2827-2831 (1973).
66. Rogers M.E., Fraser R.S.S and Loening U.E. JMB 49 681-692 (1970).
67. Grierson D., Rogers M.E., Sartirana M.L. and Loening U.E. Cold Spring Harbor Symp.Quant.Biol. 35 589-598 (1970).
68. Gall J.G. National Cancer Institute Monograph 23 475-488 (1966).
69. Rogers M.E. and Klein G. BJ 130 281-288 (1972).
70. Maden B.E.H., Salim M. and Summers D.F. NNB 237 5-9 (1972).
71. Salim M. and Maden B.E.H. Nature 244 334-336 (1973).
72. Vaughan M.H., Soeiro R., Warner J.R. and Darnell J.E. PNAS 58 1527-1534 (1967).
73. Perry R.P. and Kelley D.E. JMB 70 265-279 (1972).
74. Mirault M.E. and Scherrer K. FEBS Letters 20 233-238 (1972).
75. Prestakyo A.W., Lewis B.C. and Busch H. BBA 269 90-103 (1972).
76. Perry R.P. Biochemistry (1974) in press.

77. Meselson M., Yuan R. and Heywood J. ARB 41 447-466 (1972).
78. Loening U.E., Grierson D., Rogers M.E. and Sartirana M.L. FEBS Symposium 23 395-405 (1972).
79. Grierson D. and Loening U.E. NNB 235 80-82 (1972).
80. Tiollais P., Galibert F. and Boiron M. PNAS 68 1117-1120 (1971).
81. Choi Y.C. and Busch H. JBC 245 1954-1961 (1970).
82. Siev M., Weinberg R. and Penman S. JCB 41 510-520 (1969)
83. Hecht R.M. and Birnsteil M.L. EJB 29 489-499 (1972).
84. Parish J.H. and Kirby K.S. BBA 129 554-562 (1966).
85. Martin R.G. and Ames B.N. JBC 236 1372-1379 (1961).
86. Mahler H.R. and Cordes E.H. "Biological Chemistry" (Harper and Row, N.Y.) chap.4 (1966).
87. "Handbook of Biochemistry" ed. Sober H.A. (Chemical Rubber Co., Cleveland, Ohio) J248-251 (1968).
88. Birnie G.D. and Harvey D.R. Anal. Biochem. 22 171-174 (1968).
89. Darnell J.E., Wall R. and Tushinski R.J. PNAS 68 1321-1325 (1971).
90. Lee S.Y., Mendecki J. and Brawerman G. PNAS 68 1331-1335 (1971).
91. Edmonds M., Vaughan M.H. and Nakazato H. PNAS 68 1336-1340 (1971).
92. Edmonds M. and Caramela M.G. JBC 244 1314-1319 (1969).
93. Perry R.P. Biochem.Soc Symp. 37 105-116 (1973).
94. Helmkamp G.K. and Ts'o P.O.P. J.Amer.Chem.Soc. 83 138-142 (1961).
95. Spirin A.S. PNAR 1 301-346 (1963).
96. Loening U.E. BJ 113 131-138 (1969).
97. Cox R.A. BJ 98 841-857 (1966).
98. Cox R.A. and Katchalsky A. BJ 126 1039-1054 (1972).
99. Revzin A., Neumann E. and Katchalsky A. JMB 79 95-114 (1973).
100. Fresco J.R., Alberts B.M. and Doty P. Nature 188 98-101 (1960).

101. Cohn W.E. ME 3 724-743 (1957).
102. Bock R.M. ME 12A 224-228 (1967).
103. Tener G.M. ME 12A 398-404 (1967).
104. Markham R., Matthews R.E.F. and Smith J.D. Nature 173 537-539 (1954).
105. Singh H. and Lane B.G. Canad.J.Biochem. 42 1011-1021 (1964).
106. Lane B.G. and Tamoaki T. JMB 27 335-348 (1967).
107. Rushizsky G.W. and Sober H.A. BBA 52 217-220 (1962).
108. Cohn W.E. in "The Nucleic Acids" 1 ed. Chargaff E. and Davidson J.N. (Academic Press) 211-241 (1955).
109. Randerath K. and Randerath E. ME 12A 323-347 (1967).
110. Southern E.M. and Mitchell A.R. BJ 123 613-617 (1971).
111. Jorgensen S.E., Buch L.B. and Nierlich D.P. Science 164 1067-1070 (1969).
112. Roblin R. JMB 36 125-135 (1968)
113. Ryskov A.P. and Georgiev G.P. FEBS Letters 8 186-188 (1970).
114. Soave C., Nucca R., Sala E., Viotti A. and Galante E. EJB 32 392-400 (1973).
115. Tomlinson R.V. and Tener G.M. Biochemistry 2 703-706 (1963).
116. Takanami M. JMB 23 135-148 (1967).
117. Sugiura M. and Takanami M. PNAS 58 1595-1602 (1967).
118. Wimmer E., Chang A.Y., Clark J.M. and Reichmann M.E. JMB 38 59-73 (1968).
119. Suzuki J. and Haselkorn R. JMB 36 47-56 (1968).
120. deWachter R. and Fiers W. Nature 221 233-235 (1969).
121. Bird A., Rogers E. and Birnsteil M. NNB 242 226-230 (1973).
122. Scheer U., Trendelenberg M.F. and Franke W.W. ECR 80 175-190 (1973).

123. Grummt I. and Lindigkeit R. EJB 36 244-249 (1973).
124. Barnard E.A. ARB 38 677-732 (1969).
125. Nazar R.N. and Busch H. JBC 249 919-929 (1974).
126. Sanger F., Brownlee G.G. and Barrell B.G. JMB 13 373-398 (1965).
127. Wetmur J.G and Davidson N. JMB 31 349-370 (1968).
128. Hall J.B. and Sinsheimer R.L. JMB 6 115-127 (1963).
129. Fiers W., Lepyotre L. and Vandreissche L. JMB 13 432-450 (1965).
130. Matthews H.R. J.Gen.Virol. 3 403-415 (1968).
131. Wilkinson G.N. BJ 80 324-332 (1961).
132. Feller W. "An Introduction to Probability Theory and its Applications" (John Wiley & Sons Inc.) Vol.I chap.4 (1950).
133. Southern E.M. NNB 232 82-83 (1971).
134. Sutton W.D. and McCallum M. NNB 232 83-85 (1971).
135. Loening U.E. Biochem.Soc.Symp. 37 95-104 (1973).
136. Margoliash E., Fitch W.M. and Dickerson R.E. Brookhaven Symposium Vol.II 259-305 (1968).
137. Richardson C.C. PNAS 54 158-169 (1965)
138. Hughes S.G. and Brown P.R. BJ 131 583 (1973).
139. "Basic Phytobiophysics Training Program" California Institute of Technology Division of Biology (1968).

Eur. J. Biochem. *43*, 59–67 (1974)

5′-Ends of Ribosomal and Ribosomal Precursor RNAs from *Xenopus laevis*

Jonathan M. W. Slack and Ulrich E. Loening

Department of Zoology, University of Edinburgh

(Received August 14/October 8, 1973)

The 5′-ends of 18-S, 28-S and 40-S RNA from *Xenopus laevis* have been studied by two techniques. The first involves fractionation of alkaline hydrolysates on columns of DEAE-Sephadex at pH 7.8 and 2.7, followed by thin-layer chromatography on polyethyleneimine cellulose. The second involves labelling *in vitro* of the RNA species at the 5′-end using polynucleotide kinase.

The 40-S molecule is shown not to bear a terminal 5′-triphosphate. It has the 5′-terminus guanosine 5′-phosphate. Both the 18 and 28-S RNAs have the 5′-terminus uridine 5′-phosphate. The 28-S RNA has the 5′-terminal sequence pU-N-N-N-Gp.

These results are discussed in relation to the transcription of the ribosomal genes, the possible heterogeneity of the ribosomal precursor and the position of the 28-S sequence within the precursor molecule.

A great deal is now known about the structure of the ribosomal genes of *Xenopus laevis* [1]. In contrast, most of what is known of the synthesis, structure and processing of ribosomal RNA comes from studies on mammalian cells [2,3]. Since there are some major differences between rRNA synthesis in mammals and *Xenopus* [4,5] it is desirable to know more about the structure of the *Xenopus* ribosomal RNA species in order to increase our knowledge of the relationship between the gene and the gene product.

In this paper we address ourselves to three questions. First; is the 40-S ribosomal precursor molecules the primary product of transcription? If it is then we would expect it to bear a triphosphate group on its 5′-terminus. The method of investigation that we have adopted is to search the alkaline hydrolysis products for molecules of the type pppXp.

Secondly, previous experiments in our laboratory have demonstrated that the 40-S molecule shows heterogeneous behaviour on polyacrylamide gels [6]. One possible explanation for this is that the initiation of synthesis does not start at exactly the same point in each of the repeated gene copies. If this were the case we might expect to find more than one 5′-terminal nucleotide, and so we have identified the 5′-termini of the 18, 28 and 40-S RNAs by the method of alkaline hydrolysis.

Finally, recent evidence from electron microscopy [7] indicates that the 28-S sequence lies at the proximal end of the 40-S molecule. We have examined oligonucleotides formed by T1 RNAase digestion of 40 and 28-S RNA terminally labelled with polynucleotide kinase in order to find whether or not the 28-S sequence is at the extreme 5′-end. This appears to be the case in Novikoff cell RNA where a 5′-sequence pCmpUp was obtained from 45-S, 35-S and 28-S nucleolar RNAs [8].

Abbreviations. DEAE, diethylaminoethyl; PEI, polyethyleneimine; Butyl PBD, 2-(4′-*tert*-butylphenyl)-5-(4″-biphenylyl)-1,3,4-oxadiazole.

Enzymes. Alkaline phosphatase (EC 3.1.3.1); polynucleotide kinase (EC 2.7.1.–); pancreatic ribonuclease (EC 2.7.7.16); T1 ribonuclease (EC 2.7.7.26).

Definition. 1 unit of polynucleotide kinase catalyses the transfer of 1 nmol phosphate from ATP to CpU in 30 min at 37 °C.

MATERIALS AND METHODS

Growth and Labelling of Cells

Xenopus kidney cells were grown in Winchester bottles on a modified Matburn rolling apparatus. The growth medium consisted of minimal Eagle's medium (Glasgow modification) plus 1X non-essential amino acids, glutamine, 10% foetal bovine serum, 50 units/ml penicillin + streptomycin (all solutions from Biocult Laboratories, Glasgow) and an additional 16% of distilled water. The cells grew at room temperature in an atmosphere of 5% CO_2.

For the preparation of ribosomal precursor RNA, 2 μg/ml cycloheximide were added 2 h before extraction. We have shown that the inhibitor causes an

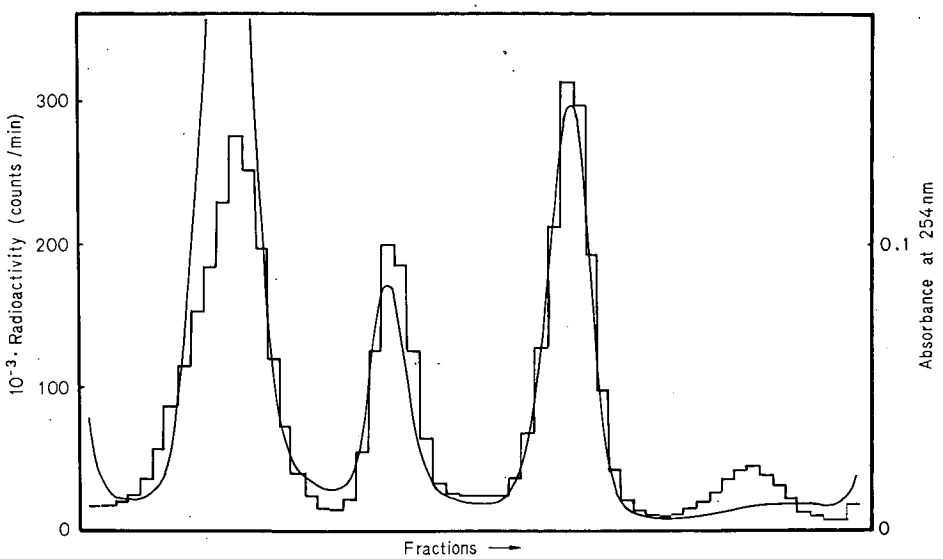


Fig. 1. *Fractionation of a total RNA preparation in the zonal rotor.* Conditions are as described in the text. This preparation has been labelled with ^{32}P for 24 h. It was DNAase-treated for 1 min so the DNA appears in the first peak. Continuous line shows the absorbance trace

approximately two-fold accumulation of precursor RNA in these cells.

Labelling with ^{32}P was carried out for 1 h to make total pulse-labelled RNA, for 3 h to make ribosomal precursor RNA and for 24 h to make 18 and 28-S RNA. For each Winchester 2—3 mCi $^{32}P_i$ (80 Ci/mg *P*, Radiochemical Centre, Amersham) were used in 50 ml phosphate-free medium containing 5% (undialysed) serum.

Preparation of RNA Species

The cells were washed in Dulbecco's phosphate-buffered saline and then dissolved in 50 mM Tris-Cl pH 7.4, 1% NaCl, 2% sodium triisopropylnaphthalenesulphonate. The method of deproteinization, removal of phenol and DNAase treatment is described elsewhere [4]. Some preparations were treated for only 1 min with DNAase instead of the usual 15 min. This causes the DNA to be broken to small pieces and it sediments with the soluble RNA in the first peak off the zonal rotor.

The total RNA preparations were fractionated into four peaks in the MSE B XIV (titanium) zonal rotor (Fig. 1). The gradient consisted of 550 ml 5—20% sucrose calculated to be linear with radius inside the rotor and hence approximately isokinetic. The buffer was 50 mM Tris-Cl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% sodium dodecylsulphate. The sample was applied on top of the gradient in 10 ml buffer + 2.5% sucrose and was overlain with 100 ml buffer. Centrifugation in the MSE Superspeed 65 centrifuge was at 45000 rev./min for 3 h 10 min at 20 °C. 10-ml fractions were collected from the rotor and for labelled preparations these were counted for 1 min by Cerenkov radiation in a Packard model 3320 scintillation counter (gain 55%, full window). Each of the four peaks was pooled and precipitated with two volumes of ethanol.

The high resolution afforded by the zonal rotor means that the two central peaks of 18 and 28-S RNA require no further purification. The 40-S peak still contains some contaminating 28-S RNA and for the alkaline hydrolysis experiments it was purified by electrophoresis for 4 h on 2.4% polyacrylamide gels. The making, running, scanning and slicing of these gels has been described previously [9]. For this and other occasions on which RNA was extracted from a gel and used for further experiments the following extraction procedure was used: the relevant gel slices were extracted for several hours with 0.6 M lithium acetate, 0.5% sodium dodecylsulphate and the RNA was pelleted from the solution by overnight centrifugation at 40000 rev./min.

Thin-Layer Chromatography

This technique is referred to several times in the following text. It was carried out on plastic-backed sheets of cellulose impregnated with polyethylene imine (obtained from Camlab, Cambridge) [10, 11]. Before use these were washed in distilled water, 10% NaCl, distilled water and 2 M pyridine formate pH 2.2. All chromatography was in an upwards direction, the samples being applied in a few microlitres of distilled water and accompanied by a marker dye mixture consisting of xylene cymol FF, orange G and acid fuschin. After running, the radioactive spots were visualised by autoradiography.

Enzyme Digestions

All enzyme digestions referred to subsequently were carried out in small drops of solution on glass-backed polythene sheets. The drops were protected from evaporation by plastic caps stuck down with silicone grease and were incubated in a 37 °C oven.

Analysis of Alkaline Hydrolysates of RNA

Alkaline Hydrolysis. The RNA samples were hydrolysed in 1—2 ml 0.2 M KOH for 24 h at 37 °C. The hydrolysates were titrated to neutrality with 5% perchloric acid and the precipitates of $KClO_4$ removed by centrifugation.

Column Chromatography. The initial fractionation was on columns (3×250 mm) of DEAE-Sephadex. The buffer was 20 mM Tris-Cl pH 7.8, 7 M urea and the salt gradient was 100 ml from 0—0.4 M NaCl [12]. 0.1 mg yeast RNA digested with pancreatic RNAase was used as an optical marker. This system gives a regular pattern of isostich peaks each with one more negative charge than the one before. The *n*th peak has a negative charge of $n + 1$ and so a radioactive peak eluting with the *n*th optical peak is referred to as the $-(n + 1)$ peak.

Some of the peaks of radioactivity were further fractionated on similar columns run at acid pH. The buffer was 7 M urea titrated to pH 2.7 with HCl and the gradient was 0—80 mM NaCl [12].

The 1-ml fractions were counted by Cerenkov radiation as described above. The recovery of counts was about 100% for the pH 7.8 columns and about 80% for the pH 2.7 columns.

Removal of Salt and Urea. This was carried out by a procedure based on that of Rushizky and Sober [13]. The pooled fractions were diluted four times with distilled water and absorbed onto 1-ml columns of DEAE-cellulose equilibrated in 20 mM ammonium carbonate pH 8. These were washed with 30 ml of the same buffer and the radioactive material eluted with 30% triethylammonium carbonate pH 9.7. The volatile salt was removed by repeatedly dissolving the sample in distilled water and evaporating to dryness on a polythene sheet with a vacuum desiccator.

Alkaline Phosphatase Digestion. Samples were digested with 0.5 μg enzyme (Whatman Biochemicals, Maidstone, England) in 5 μl 20 mM triethylammonium carbonate pH 8, 5 mM $MgCl_2$ at 37 °C for 30 min. After this the solution was evaporated and the sample redissolved in distilled water.

Thin-Layer Chromatography. Samples were run in 1 ml pyridine formate pH 3.5 or 2 M sodium formate pH 3.4. These solvents have similar properties but the second will separate all the nucleoside 3′,5′-bisphosphates while the first does not separate pAp from pCp. A mixture of radioactive standards was prepared by labelling an RNA hydrolysate with polynucleotide kinase and [γ-^{32}P]ATP. This contained all the nucleoside 3′,5′-bisphosphates together with P_i and ATP.

Labelling of RNA with Polynucleotide Kinase in vitro

Preparation of Polynucleotide Kinase. The enzyme was prepared from 50 g *Escherichia coli* B/r/T1 infected with T4 am82 phage by the procedure of Hughes and Brown [14]. The method of preparation was modified from that of Richardson [15]. The enzyme was stored frozen in 10% dimethyl sulphoxide, under which conditions we found it to be quite stable.

Alkaline Phosphatase Treatment. Purified, unlabelled RNA species were treated with half their weight of alkaline phosphatase in 0.5 ml 50 mM Tris-Cl pH 7.4, 10 mM $MgCl_2$. After incubation at 37 °C for 30 min, 0.1 ml 3 M NaCl was added and the enzyme removed by phenol extraction. The RNA was twice reprecipitated with two volumes of ethanol.

5′-Terminal Labelling. The incubation mixture contained 5 μg RNA, 3 units polynucleotide kinase, 2 nmol [γ-^{32}P]ATP (15 Ci/mmol, Radiochemical Centre, Amersham) in 20 μl 50 mM Tris-Cl pH 7.4, 5 mM $MgCl_2$, 20 mM mercaptoethanol. Incubation was for 45 min at 37 °C. After this, 20 μl formamide were added to denature the RNA and the mixture run for 3—4 h on a 2.4% polyacrylamide gel. This freed the product of radioactive ATP and degradation products. The gels were sliced and counted by Cerenkov radiation and the RNA recovered from the peak slices by the method described above.

5′-End Analysis. 5′-labelled RNA samples were hydrolysed in 10% piperidine, 1 mM EDTA in sealed capillaries at 60 °C for 48 h [9]. The piperidine was removed by evaporation and the labelled nucleoside 3′,5′-bisphosphate identified by thin-layer chromatography in 1 M pyridine formate pH 3.5. Radioactive pUp was prepared from UpC (Sigma, London) by labelling with polynucleotide kinase followed by piperidine hydrolysis of the pUpC.

5′-Oligonucleotide Analysis. The pellets of RNA recovered from the gels after 5′-labelling were dissolved in distilled water and freed from salts and dodecylsulphate by passage through 1-ml columns of Sephadex G-100. The fractions containing radioactivity were dried down on a polythene sheet and incubated with 0.5 μg T1 RNAase and 5 μg unlabelled tRNA in 5 μl 10 mM Tris-Cl pH 7.6, 1 mM EDTA, for 45 min at 37 °C. The resulting 5′-labelled oligonucleotide was characterised by two dimensional chromatography on polyethyleneimine thin layers, the first dimension in 2 M pyridine formate pH 3.5 and the second in 1 M LiCl. The LiCl was removed by washing in methanol.

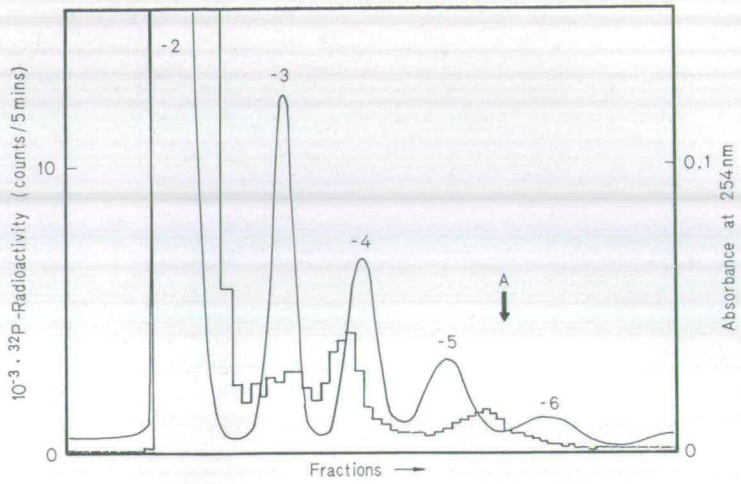


Fig. 2. *DEAE-Sephadex chromatography at pH 7.8.* The continuous line shows the absorbance trace of yeast RNA digested with pancreatic RNAase. The histogram shows ^{32}P radioactivity from alkali-hydrolysed total RNA pulse labelled for 1 h. The peak labelled A is believed to consist of molecules of the type pppNp

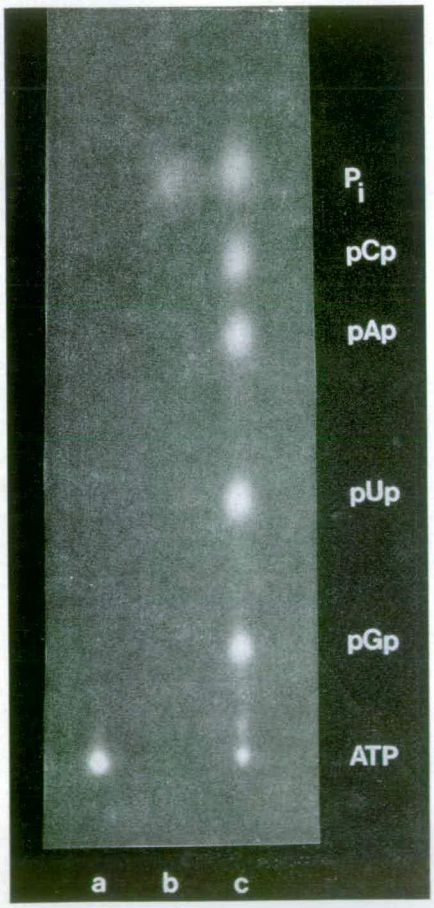


Fig. 3. *Thin-layer chromatography on PEI-cellulose.* The solvent is 2 M sodium formate pH 3.4. (a) Peak A; (b) peak A after treatment with alkaline phosphatase; (c) radioactive standards with the identity of each spot shown

RESULTS

Alkaline Hydrolysis of Total Pulse-Labelled RNA

This was a control experiment to find whether 5′-triphosphates are formed in *Xenopus* RNA *in vivo*. After alkaline hydrolysis these should be liberated as molecules of the type pppNp which have been reported to eluate from DEAE-Sephadex columns between the fourth and fifth isostich peaks [16].

The RNA used for these experiments was pulse-labelled with ^{32}P for 1 h. We found that the bulk of the radioactivity in such preparations was actually contained in small molecules such as ATP which coprecipitated with the RNA in ethanol. The RNA was freed from these contaminants by dialysis for 48 h against two changes of 0.6 M lithium acetate, 1% sodium dodecylsulphate.

The DEAE-Sephadex columns showed a small peak of radioactivity eluting between the 4th and 5th isostich and none at higher salt concentration (Fig. 2). This material was found to be immobile on PEI-cellulose and to lose all its phosphate when treated with alkaline phosphatase (Fig. 3). We conclude that molecules of the type pppNp can readily be demonstrated by these techniques providing that the starting material contains sufficient radioactivity.

Alkaline Hydrolysis of 18, 28 and 40-S RNA

The 18-S hydrolysate gave a large peak eluting with the second isostich and a small one with the third isostich (Fig. 4 A and Table 1). The —3 peak was presumed to consist of 2′-*O*-methylated dinucleotides which are resistant to alkaline hydrolysis. The —4 peak was shown to be a single component by DEAE-Sephadex chromatography at pH 2.7. It

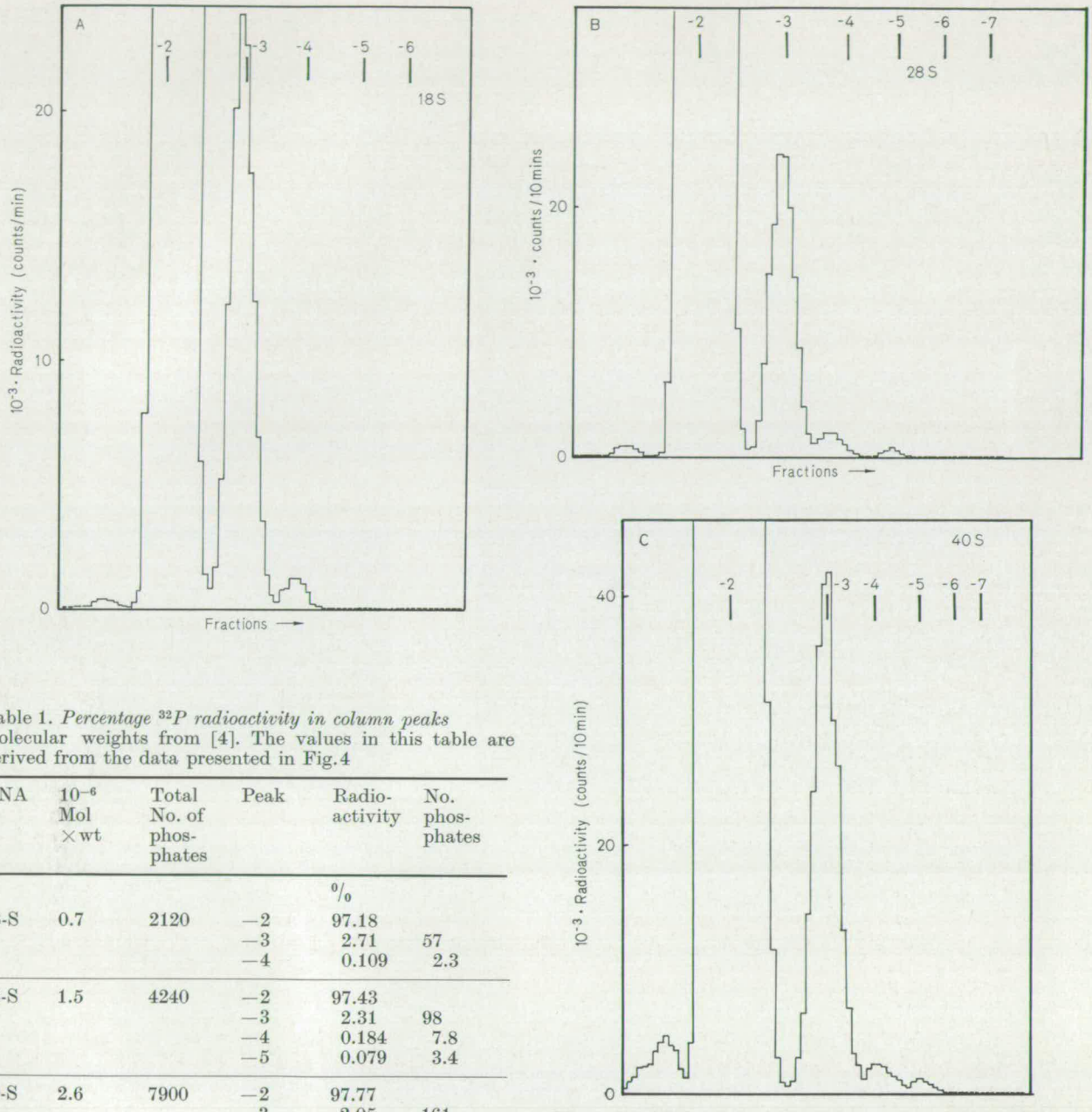


Fig. 4. *DEAE-Sephadex chromatography at pH 7.8.* The bars show the position of elution of the isostichs from yeast RNA digested with pancreatic RNAase. The numbers indicate the net negative charge of each isostich. The histograms show ^{32}P radioactivity from alkaline hydrolysates of individual species of RNA. (A) 18-S RNA; (B) 28-S RNA; (C) 40-S RNA

Table 1. *Percentage ^{32}P radioactivity in column peaks*
Molecular weights from [4]. The values in this table are derived from the data presented in Fig. 4

RNA	10^{-6} Mol ×wt	Total No. of phos-phates	Peak	Radio-activity	No. phos-phates
				%	
18-S	0.7	2120	−2	97.18	
			−3	2.71	57
			−4	0.109	2.3
28-S	1.5	4240	−2	97.43	
			−3	2.31	98
			−4	0.184	7.8
			−5	0.079	3.4
40-S	2.6	7900	−2	97.77	
			−3	2.05	161
			−4	0.125	10.2
			−5	0.055	4.5

co-chromatographed with pUp on PEI-cellulose and lost all its phosphate on treatment with alkaline phosphatase (Fig. 5). Therefore we conclude that it is pUp from the 5′-end of the 18-S molecule.

Both the 28 and 40-S RNAs gave peaks eluting with the second, third and fourth isostich (Fig. 4 B and C, Table 1). The −5 peak is in nearly the right place to represent a triphosphate end, but we present evidence in the accompanying paper that it is actually a tetranucleotide with three adjacent 2′-*O*-methylations. The hydrolysis products from the 5′-termini were found in the −4 peaks. Chromatography of this material on DEAE-Sephadex at pH 2.7 gave three

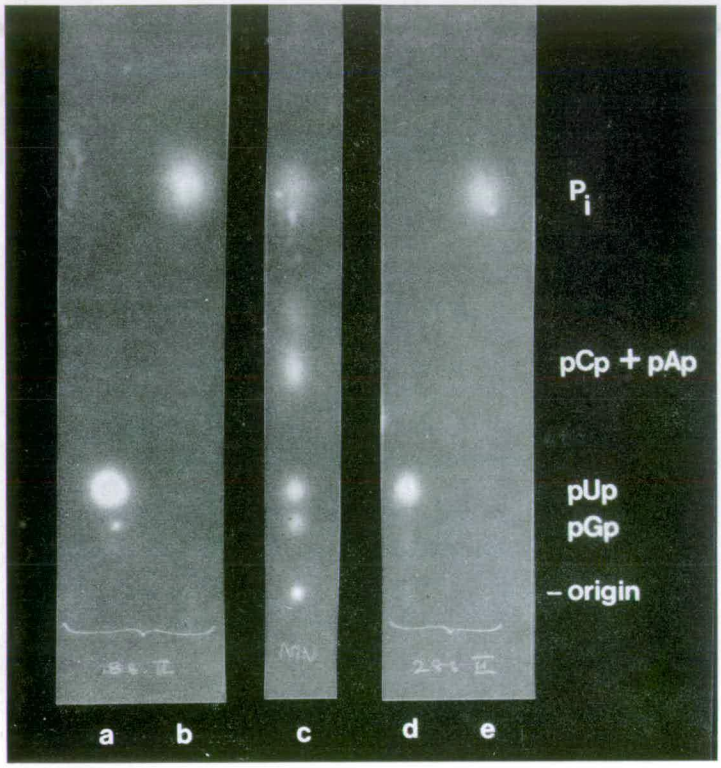


Fig. 5

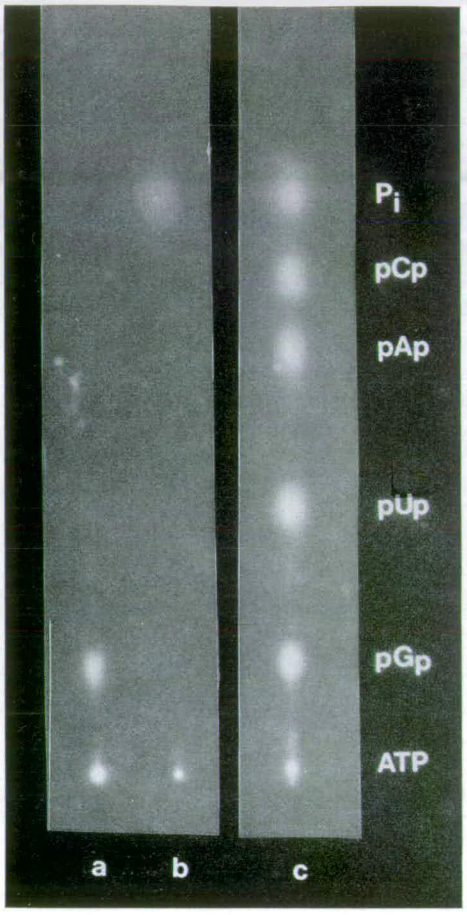


Fig. 6

Fig. 5. *Thin-layer chromatography on PEI-cellulose.* The solvent is 1 M pyridine formate pH 3.5. This shows the behaviour of the alkaline hydrolysis products from 5′-termini. (a) from 18-S RNA; (b) the same after alkaline phosphatase treatment; (c) standards with identities indicated at the side; (d) from 28-S RNA; (e) the same after alkaline phosphatase treatment

Fig. 6. *Thin-layer chromatography on PEI-cellulose.* The solvent is 2 M sodium formate pH 3.4. (a) alkaline hydrolysis product from 5′-terminus of 40-S RNA; (b) the same after alkaline phosphatase treatment; (c) standards with identities indicated

peaks. For both the 28-S and 40-S RNAs, the second of these peaks lost all its phosphate on treatment with alkaline phosphatase while the others did not. The thin-layer chromatography showed that pUp was obtained from the 28-S RNA and pGp from the 40-S RNA (Fig. 5 and 6).

5′-Termini from Labelled RNA in vitro

18 and 28-S RNA could be labelled to a high specific activity at the 5′-end by the polynucleotide kinase reaction as described in the Materials and Methods section. It was demonstrated that the products were undegraded by adding a sample of tritiated total RNA after completion of the reaction. The gels were counted for ^{32}P and ^{3}H (3:2 v/v toluene : methoxyethanol + 0.5% w/v, butyl PBD, ^{3}H gain 70%, window 50—500, ^{32}P gain 1.8%, window 50—1000), and for both RNA species the 5′-labelled RNA had the same mobility as the tritiated marker (Fig. 7 A and B).

Analysis of the 5′-termini by piperidine hydrolysis and thin-layer chromatography is shown in Fig. 8. In agreement with the results from the alkaline hydrolysis study, both species were found to give pUp.

5′-Oligonucleotides

The polynucleotide kinase method was used to label 40-S RNA prepared with the zonal rotor and treated with alkaline phosphatase (Fig. 9). A comparison of the oligonucleotides obtained from 40-S and 28-S RNA after digestion with T1 RNAase is

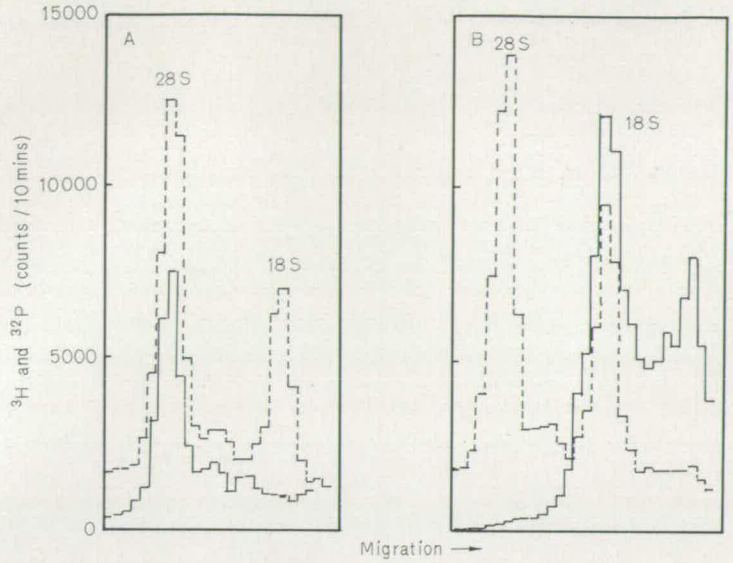


Fig. 7. *Electrophoresis on 2.4% polyacrylamide gels.* The dashed histograms represent ^{3}H radioactivity from a total RNA preparation labelled *in vivo*. The solid histograms represent ^{32}P radioactivity from individual RNA species labelled *in vitro* with polynucleotide kinase. (A) 28-S RNA; (B) 18-S RNA

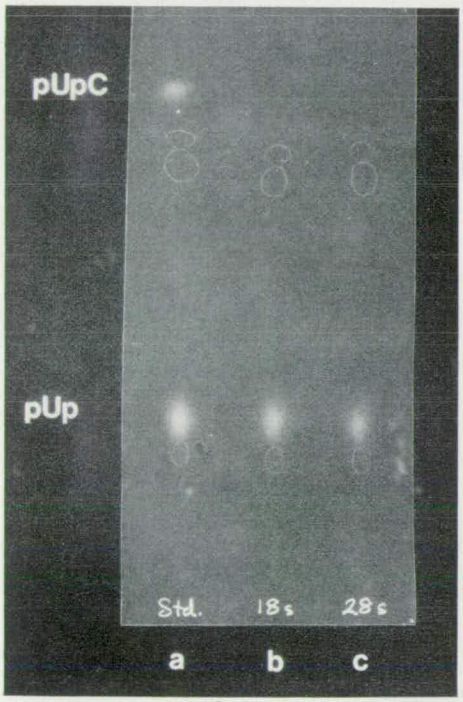


Fig. 8. *Thin-layer chromatography on PEI-cellulose.* The solvent is 1 M pyridine formate pH 3.5. (a) standards with identities indicated; (b) piperidine hydrolysate of 18-S RNA labelled with polynucleotide kinase; (c) the same for 28-S RNA

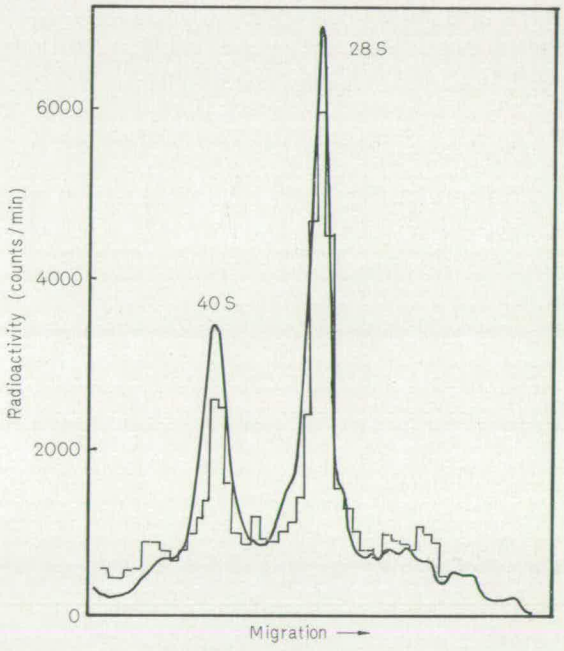


Fig. 9. *Electrophoresis on 2.4% polyacrylamide gels.* The sample is RNA from the fourth zonal rotor peak which has been treated with alkaline phosphatase and labelled at the 5′-end with polynucleotide kinase. The continuous line represents absorbance and the histogram ^{32}P radioactivity

shown in Fig. 10. The product from 40-S RNA moved in the position to be expected of pGp (just behind the orange G marker). The principal product from the 28-S RNA was rather slower. The length of this oligonucleotide was determined by chromatography on DEAE-Sephadex with unlabelled T1 RNAase digested yeast RNA as marker (Fig. 11). The radioactive peak eluted with the seventh isostich and

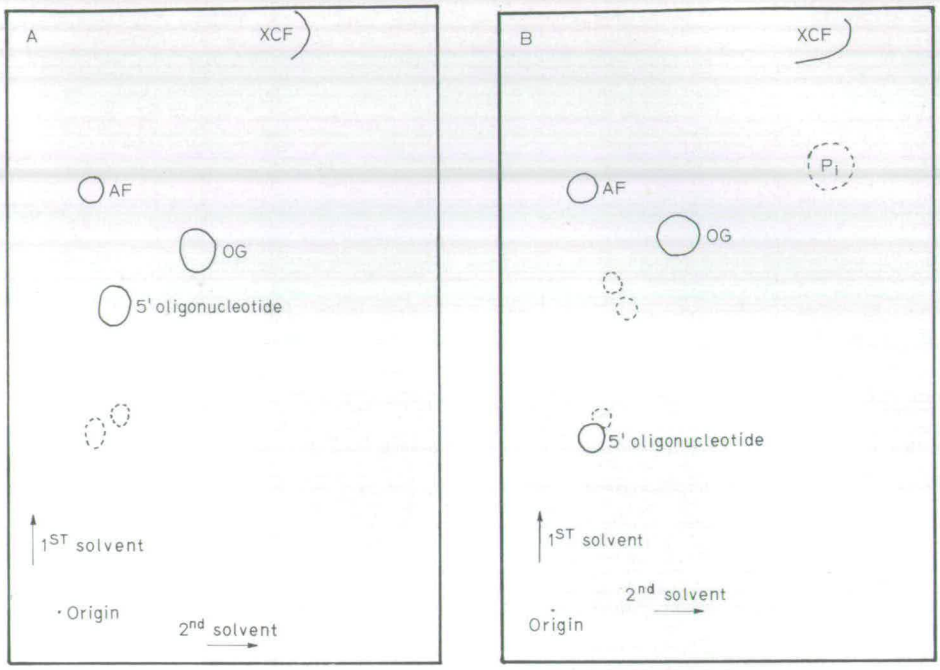


Fig. 10. *Two-dimensional chromatography on PEI-cellulose.* The first solvent is 2 M pyridine formate pH 3.5, and the second is 1 M lithium chloride. The samples are T1 RNAase digests of 5′-terminally labelled RNA. The positions of the marker dyes are indicated.: XCF xylene cymol FF, OG orange G and AF acid fuschin. (A) 40-S RNA; (B) 28-S RNA

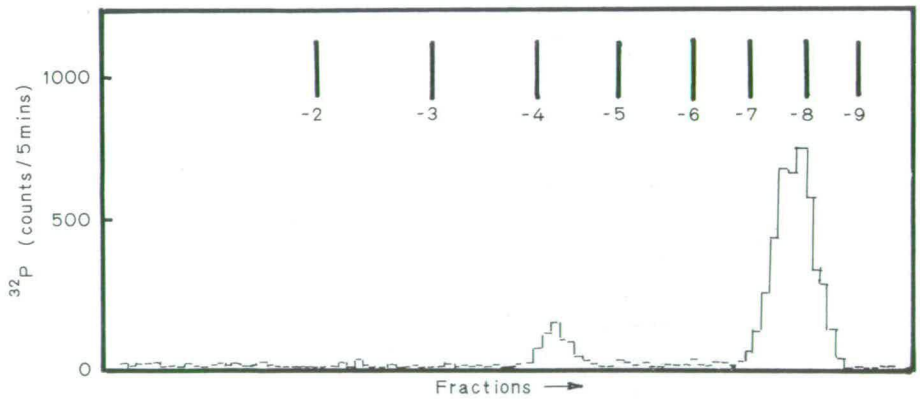


Fig. 11. *DEAE-Sephadex chromatography at pH 7.8.* The bars indicate the positions of isostichs from yeast RNA digested with T1 RNAase. The histogram represents ^{32}P radioactivity from the 5′-terminal oligonucleotide of 28-S RNA

therefore has a net charge of −8. Since the 3′-residue must be guanosine, we can conclude that it has the sequence pU-N-N-N-Gp.

DISCUSSION

The implications of our results for the questions posed in the introduction would seem to us to be as follows:

a) The 40-S molecule does not bear a 5′-triphosphate. This suggests that it may not be a primary transcription product although it is possible that the two extra phosphates are removed shortly after transcription. If there is an even larger precursor then it must be extremely short lived. A recent careful search for heavy rRNA molecules in *Xenopus* oocytes has proved negative [17]. An alternative explanation is that the mechanism of movement of RNA polymerase along the untranscribed spacer regions involves going through all the motions of transcription except forming the phosphodiester bonds. This type of movement would require a supply

of 5′-triphosphates and convert them to 5′-monophosphates and one might then expect primary transcripts to bear only a single 5′-phosphate.

b) We have shown that all three ribosomal RNA species bear a single 5′-terminal nucleotide. This makes it unlikely that the heterogeneous behaviour of 40-S RNA on polyacrylamide gels results from it being composed of a mixture of molecules of different lengths, although to be sure of this a study of the 3′-end would also be necessary.

c) Since the 5′-ends of the 40-S and 28-S RNAs are different, the processing mechanism must include at least one reaction at this end of the precursor molecule. This makes it necessary to consider more complicated processing schemes than that suggested by Perry and Kelley [18].

This work was supported by the Medical Research Council. JMWS holds the Ramsey Wright Scholarship of Edinburgh University. We would like to thank Steve Hughes and Alan Morrison of the Department of Molecular Biology for assistance with the preparation of polynucleotide kinase, Anne Baker for assistance with *Xenopus* cell culture and Dr E. M. Southern for assistance with thin-layer chromatography.

REFERENCES

1. Birnsteil, M. L., Chipcase, M. & Speirs, J. (1971) *Prog. Nucleic Acid Res. Mol. Biol. 11*, 351–389.
2. Loening, U. E. (1970) *Symp. Soc. Gen. Microbiol. 20*, 77–106.
3. Maden, B. E. H. (1971) *Prog. Biophys. 22*, 129–178.
4. Loening, U. E., Jones, K. & Birnsteil, M. L. (1969) *J. Mol. Biol. 45*, 353–366.
5. Grierson, D., Rogers, M. E., Sartirana, M.-L. & Loening, U. E. (1970) *Cold Spring Harbor Symp. Quant. Biol. 35*, 589–598.
6. Loening, U. E., Grierson, D., Rogers, M. E. & Sartirana, M.-L. (1971) in *Functional Units in Protein Biosynthesis* (Cox, R. A. & Hadjiolov, A. A., ed.) pp. 395–405, Academic Press, London.
7. Brown, D. D., Sugimoto, K. & Carroll, D. (1973) *Cold Spring Harbor Symp. Quant. Biol. (Abstracts) 38*, 50.
8. Choi, Y. C. & Busch, H. (1970) *J. Biol. Chem. 245*, 1954–1961.
9. Loening, U. E. (1967) *Biochem. J. 102*, 251–257.
10. Randerath, K. & Randerath, E. (1967) *Methods Enzymol. 12 A*, 323–347.
11. Southern, E. M. & Mitchell, A. R. (1971) *Biochem. J. 123*, 613–617.
12. Tener, G. M. (1967) *Methods Enzymol. 12 A*, 398–404.
13. Rushizsky, G. W. & Sober, H. A. (1962) *Biochim. Biophys. Acta, 52*, 217–220.
14. Hughes, S. G. & Brown, P. R. (1973) *Biochem. J. 131*, 583.
15. Richardson, C. C. (1965) *Proc. Natl. Acad. Sci. U. S. A. 54*, 158–165.
16. Roblin, R. (1968) *J. Mol. Biol. 36*, 125–135.
17. Bird, A., Rogers, E. & Birnsteil, M. (1973) *Nat. New Biol. 242*, 226–230.
18. Perry, R. P. & Kelley, D. E. (1972) *J. Mol. Biol. 70*, 265–279.

J. M. W. Slack and U. E. Loening, Department of Zoology, University of Edinburgh, West Mains Road, Edinburgh, Great Britain EH9 3JT

Eur. J. Biochem. *43*, 69—72 (1974)

28-S RNA from *Xenopus laevis* Contains a Sequence of Three Adjacent 2′-*O*-Methylations

Jonathan M. W. Slack and Ulrich E. Loening

Department of Zoology, University of Edinburgh

(Received August 14/October 8, 1973)

Evidence is presented to show that an alkaline hydrolysis product from 40-S and 28-S RNA of *Xenopus laevis* is a tetranucleotide containing three adjacent 2′-*O*-methylations.

This is the longest such sequence yet reported and is believed to be Gm-Am-Am-Ap.

About 1% of the ribose residues in ribosomal RNA are methylated in the 2′-*O* position. This renders the corresponding phosphodiester bond resistant to alkaline hydrolysis and to cyclizing ribonucleases. This property has been used to isolate methylated dinucleotides from ribosomal RNA of various organisms by alkaline hydrolysis [1—3].

In this paper we report the discovery of an alkali stable tetranucleotide in 28-S RNA from *Xenopus laevis*. We made use of RNA double labelled with ^{32}P and with [*Me*-^{3}H]methionine. The latter isotope, under the conditions used, was confined to the methyl group protons in the RNA.

This component was found during work on the 5′-end determinations reported in the accompanying paper. We have characterised the tetranucleotide in some detail because of possible confusion between it and the 5′-terminal sequence pCm-pUp found in Novikoff cell 28-S RNA [4]. This latter sequence has been used to determine the order of the 28 and 18-S sequences in the ribosomal precursor, but our experiments indicate that it does not exist in *Xenopus*.

MATERIALS AND METHODS

Most of the techniques used are the same as those described in the accompanying paper. The additional ones were as follows.

*Labelling of RNA with [*Me-^{3}H*]Methionine*

Xenopus kidney cells were labelled for 24 h in methionine-free medium containing 5% (undialysed) serum, 0.05 mCi/ml [*Me*-^{3}H]methionine (9 Ci/mmol, Radiochemical Centre, Amersham) with 20 μM adenosine, 20 μM guanosine and 20 mM sodium formate to suppress the labelling of the purine rings [5]. The relatively small amount of ^{3}H labelling in the mononucleotide peaks of Fig. 1 indicates that only the methyl group protons were labelled.

Total Digestion with Venom Phosphodiesterase

This was carried out in 5 μl 50 mM Tris-acetate pH 8.2, 5 mM $MgCl_2$ with 1 μg enzyme (Worthington Biochemical Corp., New Jersey, U.S.A.). Incubation was for 3 h at 37 °C.

Separation Methods

Paper chromatography was carried out on Whatman 3 MM paper with a solvent mixture composed of ethyl acetate—isopropanol—7.5 M ammonia—*n*-butanol (3:2:2:1, v/v/v/v) [6].

Paper electrophoresis was carried out on Whatman 52 paper impregnated with 5% pyridine acetate pH 3.5. The paper strips of about 60-cm length were run immersed in white spirit at 4 kV for 1 h.

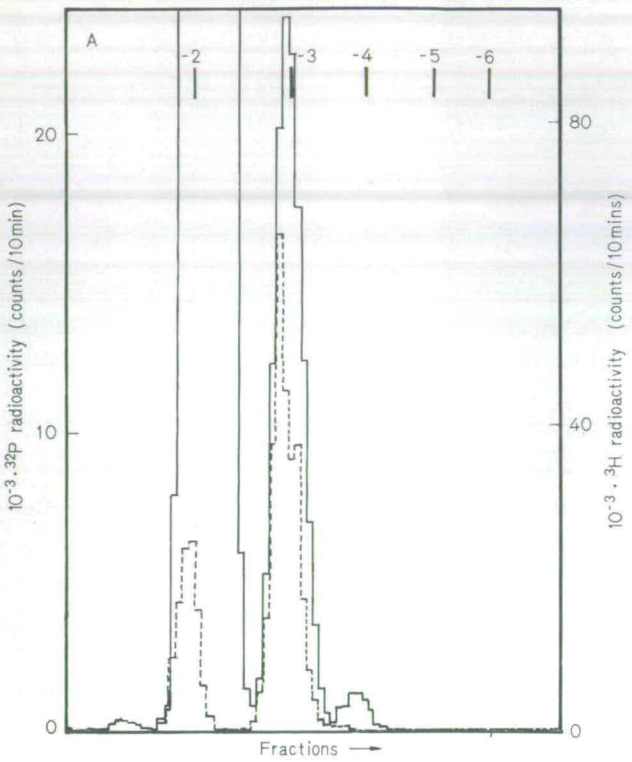
Counting Techniques

Slices of paper from chromatograms and electropherograms were counted immersed in 3:2 (v/v) toluene : methoxyethanol + 0.5% (w/v) butyl PBD in a Packard model 3320 scintillation counter. The ^{32}P channel was set at 1.8% gain and full window and the ^{3}H channel at 90% gain and 50—600 window. Under these conditions about 3% of the ^{32}P radioactivity was counted in the ^{3}H channel and this was subtracted.

Aliquots of buffer containing 7 M urea were counted in the same way. For the larger volumes it was necessary to increase the methoxyethanol content of the scintillator to 55%, which means that $^{32}P/^{3}H$ ratios cannot be compared between different experiments.

Abbreviations. DEAE, diethylaminoethyl; butyl-PBD, 2-(4′-*tert*-butylphenyl)-5-(4″-biphenylyl)-1,3,4-oxadiazole.

Enzymes. Alkaline phosphatase (EC 3.1.3.1); venom phosphodiesterase (EC 3.1.4.1).



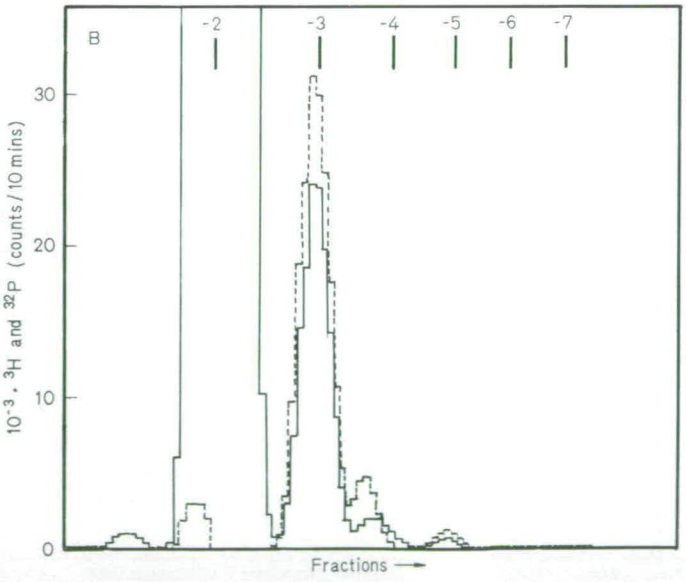


Fig. 1. *DEAE-Sephadex chromatography at pH 7.8.* The bars indicate the positions of elution of the marker isostichs and their negative charges. The continuous histogram shows ^{32}P radioactivity and the dashed histogram ^{3}H radioactivity from alkaline hydrolysates of double-labelled RNA. (A) 18-S RNA; (B) 28-S RNA

RESULTS

Alkaline-Hydrolysis of Double-Labelled RNA

In the accompanying paper it is shown that when alkaline hydrolysates of RNA are fractionated on DEAE-Sephadex columns with 7 M urea, a −5 peak is obtained from 28-S and 40-S but not from 18-S RNA. To find which of the peaks contained methylated oligonucleotides, we repeated the experiments with double-labelled RNA.

RNA labelled with [^{3}H]methionine was mixed with RNA labelled with ^{32}P and fractionated in the zonal rotor. The 18 and 28-S RNAs were recovered, hydrolysed in KOH and run on DEAE-Sephadex columns at pH 7.8. In Fig. 1 it is shown that the ^{3}H labelling patterns differ for the two RNA species. For both, the −3 peak is labelled with ^{3}H and was identified as 2′-*O*-methylated dinucleotides. For the 18-S RNA the −4 peak is unlabelled with ^{3}H which is consistent with our finding that it is pUp from the 5′-end of the molecule. But for the 28-S RNA, both the −4 and −5 peaks are labelled with ^{3}H. The ^{3}H/^{32}P ratios are shown in Table 1. The theoretical values are calculated on the assumption that the −4 peak consists of two methylated trinucleotides + pUp and the −5 peak of a single tetranucleotide.

When the material from the 28-S −4 peak was run on DEAE-Sephadex at pH 2.7, three peaks were obtained of which the first and third were labelled with tritium (Fig. 2). We have identified the centre peak as pUp from the 5′-end.

Analysis of the −5 Peak

When the material from the −5 peak was rehydrolysed for 24 h in KOH and then rerun on the DEAE-Sephadex column, it still eluted in the same position (Fig. 3). This shows that it is not an artifact of incomplete hydrolysis.

When it was subjected to treatment with alkaline phosphatase and rerun on the same column, it eluted with the second marker isostich instead of the fourth *i.e.* it has lost two negative charges (Fig. 4). This is the "shift-out" procedure of Tomlison and Tener [7]

Table 1. *Isotope ratios for 28-S RNA hydrolysis products*
These values are derived from the data presented in Fig. 1. In the third column the measured isotope ratios are divided by 2.02. The numbers in the fourth column are the ratios of the numbers of methyl to phosphate groups in the molecules believed to be present

Peak	^{3}H/^{32}P	Relative values	Theoretical
−3	1.21	0.6	0.5
−4	1.34	0.66	0.66
−5	1.53	0.75	0.75

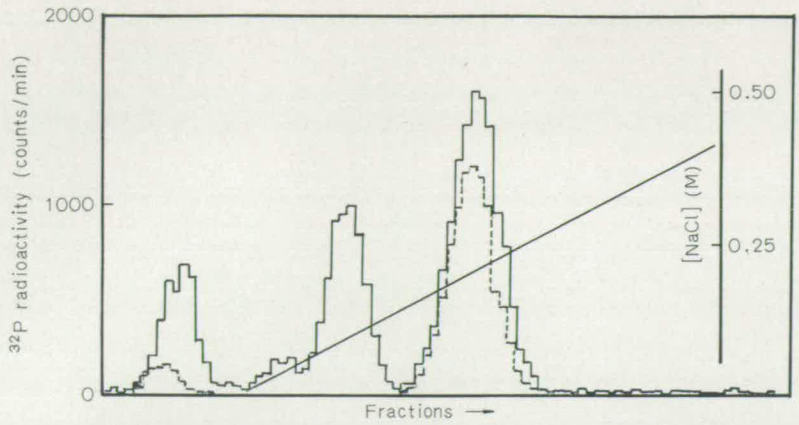


Fig. 2. *DEAE-Sephadex chromatography at pH 2.7.* The sample is the −4 peak from the 28-S RNA hydrolysate. The continuous histogram represents ^{32}P radioactivity and the dashed histogram ^{3}H radioactivity

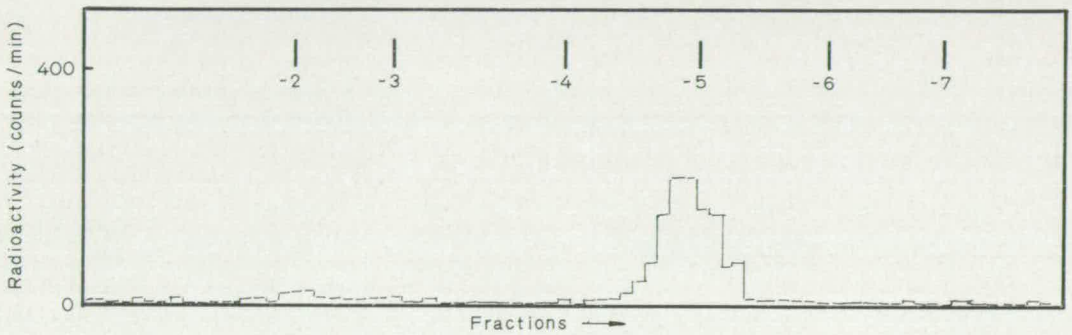


Fig. 3. *Rechromatography on DEAE-Sephadex at pH 7.8 of the −5 peak from the 28-S RNA hydrolysate after a second round of hydrolysis*

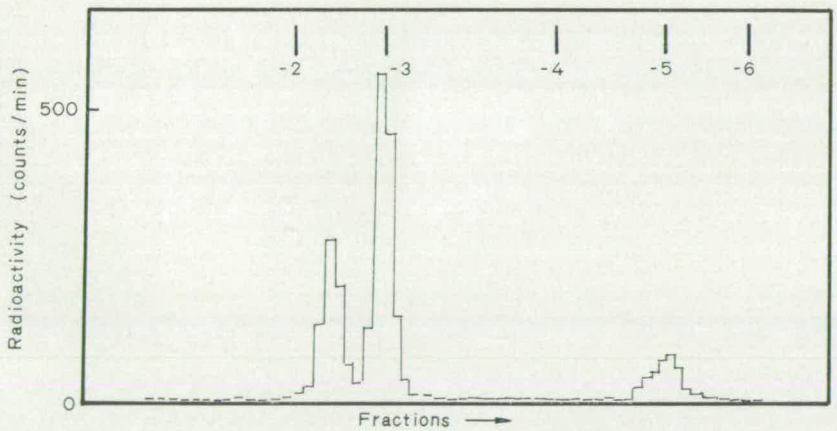


Fig. 4. *Rechromatography on DEAE-Sephadex at pH 7.8 of the −5 peak from the 28-S RNA hydrolysate after treatment with alkaline phosphatase.* Phosphatase treatment was incomplete and so some radioactivity remains in the −5 position

which serves to distinguish internal oligonucleotides (which lose two negative charges) from terminal ones (which lose 0, 4 or more).

$$\underset{-5}{\text{N-N-N-Np}} \rightarrow \underset{-3}{\text{N-N-N-N}_{OH}} + \underset{-2}{P_i}$$

For analysis of the base composition of 2′-*O*-methylated sequence it is necessary to remove the 3′-terminal phosphate with alkaline phosphatase and then to digest with venom phosphodiesterase. When the oligonucleotide is double labelled the products are one ^{3}H-labelled 2′-*O*-methyl nucleoside

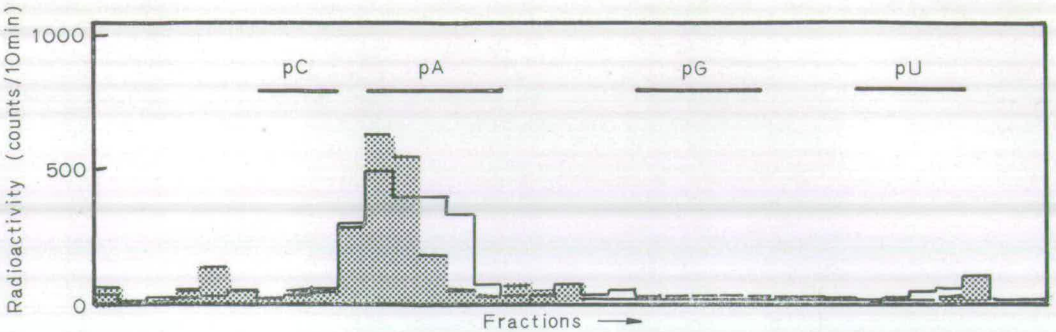


Fig. 5. *Paper electrophoresis of a venom phosphodiesterase digest of double-labelled tetranucleotide.* The solid line represents ^{32}P radioactivity and the shaded histogram represents ^{3}H radioactivity

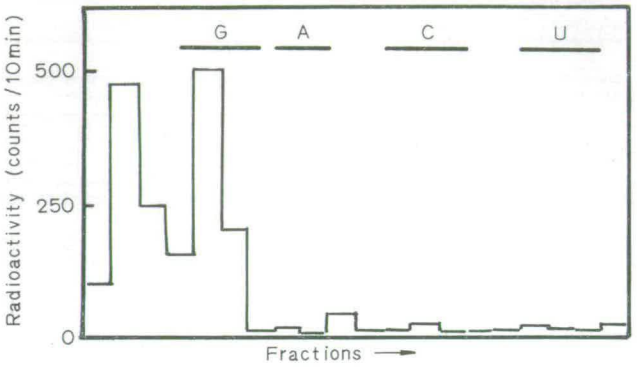


Fig. 6. *Paper chromatography of a venom phosphodiesterase digest of the tetranucleotide labelled with [*Me-^{3}H*]methionine*

from the 5′-end, one ^{32}P-labelled 5′-nucleotide from the 3′-end and two double-labelled 2′-*O*-methyl 5-nucleotides.

Such a venom phosphodiesterase digest was mixed with unlabelled 5′-nucleotides and fractionated by paper electrophoresis as described in Materials and Methods. This yielded a large radioactive peak of ^{3}H and ^{32}P running slightly behind adenosine 5′-phosphate (Fig. 5), which suggests that the three distal bases are all adenines. The small peak of ^{3}H at the origin presumably represents the 2′-*O*-methyl nucleoside.

The 2′-*O*-methyl nucleoside was identified by fractionating a venom phosphodiesterase digest by paper chromatography. (Fig. 6). With the system described, all the 5′-nucleotides move slower than the slowest nucleoside. The only nucleoside spot which was labelled was guanosine, so it was concluded that the 5′-residue is 2′-*O*-methyl-guanosine and the complete structure of the tetranucleotide is Gm-Am-Am-Ap.

DISCUSSION

If we take 1.5% as the proportion of methylated sugars in 28-S RNA then we can calculate from Poisson statistics the chance of obtained three adjacent methylations by random processes. This comes out at 1.56×10^{-5} for the whole 28-S sequence.

This suggests that the sequence has some specific function, perhaps concerned with the assembly of the larger ribosomal subunit. A recent finding (B. E. H. Maden, personal communication) that HeLa cell 28-S RNA may contain a similar trimethylated sequence strengthens this possibility. The work of Vaughan *et al.* [8] showed for HeLa cells that the methylation was necessary for ribosome assembly but not for the initial processing of the ribosomal precursor. Beyond this we know very little at present.

REFERENCES

1. Singh, H. & Lane, B. G. (1964) *Can. J. Biochem. 42*, 1011–1021.
2. Lane, B. G. & Tamoaki, T. (1967) *J. Mol. Biol. 27*, 335–348.
3. Wagner, E. K., Penman, S. & Ingam, V. (1967) *J. Mol. Biol. 29*, 371–287.
4. Choi, Y. C. & Busch, H. (1970) *J. Biol. Chem. 245*, 1954–1961.
5. Maden, B. E. H. (1972) *Nat. New Biol. 237*, 5–9.
6. Randerath, K. & Randerath, E. (1968) *Anal. Biochem. 28*, 110–118.
7. Tomlinson, R. V. & Tener, G. M. (1963) *Biochemistry, 2*, 703–706.
8. Vaughan, M. H., Soeiro, R., Warner, J. R. & Darnell, J. E. (1967) *Proc. Natl. Acad. Sci. U. S. A. 58*, 1527–1534.

J. M. W. Slack and U. E. Loening, Department of Zoology, University of Edinburgh, West Mains Road, Edinburgh, Great Britain EH9 3JT