1976

Studies of Plant Viroid RNA and Other RNA Species of Unusual Function

Elizabeth Dickson

STUDIES OF PLANT VIROID RNA AND OTHER RNA SPECIES OF UNUSUAL FUNCTION

A thesis submitted to the Faculty of The Rockefeller University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

by

Elizabeth Dickson III

April 6, 1976

The Rockefeller University

New York, New York

DEDICATION

This thesis is dedicated to my parents, David and Rosaleen Dickson, for their constant encouragement, moral support and inspiring example of life at its fullest.

PREFACE

The work described in this thesis was carried out while I held a graduate fellowship at The Rockefeller University. I am grateful to President Frederick Seitz, Vice-President Rodney W. Nichols, Dr. James G. Hirsch, Dr. Clarence M. Connelly, and the members of the administration who have created a university where research can be carried out under optimum conditions. During the final two years of study, I was the recipient of an IBM Fellowship for which I am most appreciative.

I would like to thank my thesis advisor, Dr. Hugh D. Robertson, for a totally unique experience in scientific collaboration both at the experimental and theoretical level and for his concern, advice and humor throughout my graduate career.

I would like to express special thanks to three "outside advisors" with whom I collaborated and from whom I gained valuable experience: Dr. Wolf Prensky of the Sloan-Kettering Institute, New York; Dr. Theodor O. Diener of the U.S.D.A., Beltsville, Maryland; and Dr. John J. Dunn of Brookhaven National Laboratory, Long Island, New York.

I am grateful also to all of the members of the Zinder Laboratory for making my stay at Rockefeller enjoyable and for their constant readiness to help in many ways. In particular I am indebted to Dr. Norton D. Zinder for providing an example of thinking at its best and to Dr. Peter Model for invaluable assistance with ideas and experiments.

I would also like to thank the members of Dr. Wolf Prensky's laboratory, John Ernst, Janet Robertson and Paul Szabo, for their active collaboration in studies of iodinated RNA and their aid and advice in numerous other areas.

Many other Rockefeller University Faculty members have contributed substantially to my education: Dr. Gunther Blobel, Dr. Armin C.

Braun, Dr. Bruce Cunningham, Dr. Gerald M. Edelman, Dr. Warren Jelinek, Dr. George Pieczenik, Dr. Morris Schreiber, Dr. Samuel Silverstein, Dr. Patricia Spear, Dr. William Trager and Dr. Robert Turgeon.

I would like to express my sincere thanks to Edward Pelle and Susan Bornstein for excellent technical assistance and to Louise Pape for secretarial help far beyond the call of duty in the preparation of manuscripts and this thesis.

Finally I would like to thank all those at Rockefeller who helped to make my life here easier to live: in particular Henrik Boudakian, Rudolf K. Franz, Frances Kralick, Bonnie Ann Lobue, Ray Martin, Sandra O'Brien, Virginia Sides, Regina Titus, Sandra M. Walsh and Leah Woerner to name just a few.

CONVENTIONS AND ABBREVIATIONS

DEAE-cellulose -- O-(diethylaminoethyl) cellulose

EDTA -- ethylenediaminetetraacetic acid

Tris -- tris(hydroxymethyl)amino-methane

fingerpint -- This term is defined here as a two-dimensional fractionation of oligonucleotides, usually following digestion of some sort, and refers literally to the supporting medium bearing the fractionated oligonucleotides (paper or thin-layer plate). By extension, however, the term is also applied both to autoradiographs of patterns taken on X-ray films, and to photographs of such X-ray films; fingerprint can be used as either noun or verb.

fingerprinting -- an alternative form of the word fingerprint which can be used either as noun or adjective.

CMCT -- N-cyclohexyl-N'-(β-morpholinyl-(4)-ethyl) carbodiimide-methyl-p-toluene sulfonate

5' side means left when referring to RNA sequences or drawings thereof, which are conventionally depicted with the 5' end to the left

3' side means right in this context

spot -- When an X-ray film is exposed to a fingerprint and subsequently developed, a pattern of black spots appears on the film. Each such visible region of silver grains on the film is called a spot. Furthermore, the analogous regions of the thin-layer plate containing radioactive material which caused the spots on the film are also referred to as spots. Thus it is possible to talk about eluting spots from a fingerprint.

PSTV -- potato spindle tuber viroid

CEV -- citrus exocortis viroid

CSV -- chrysanthemum stunt viroid

CPFV -- cucumber pale fruit viroid

ChCMV -- chrysanthemum chlorotic mottle viroid

SUMMARY

This dissertation concerns studies of plant viroid RNA and other RNA species of unusual function. By "RNA species of unusual function" we mean RNA species which do not function as transfer, ribosomal or messenger RNA. In the studies to be reported here we have laid the experimental and theoretical foundation for investigating the role of RNA in the control of gene expression.

In Chapter I the various standard techniques of RNA fingerprinting and sequencing have been applied to RNA molecules labeled in vitro with ^{125}I. Fingerprints of human 5S RNA and bacteriophage f2 RNA resemble those of their noniodinated counterparts both in complexity and in specific pattern. Iodination as used here is thus a general labeling procedure, and appears principally to label cytidine residues. This iodination method shows little sensitivity to potential structure in single-stranded RNA molecules, yields stable oligonucleotide products in a reproducible manner, and does not change the specificity of RNases T1, T2, and U2, nor does it affect the specificity of digestion by pancreatic RNase or spleen phosphodiesterase. Some difficulties arising in the handling of iodinated RNA are also discussed.

In Chapter II we have employed these techniques to study several properties of plant viroid RNA which was previously unobtainable as a radioactively labeled species of high specific activity. Highly purified isolates of potato spindle tuber viroid (PSTV) from tomatoes and citrus exocortis viroid (CEV) from Gynura aurantiaca have been compared. Two-dimensional fingerprinting analysis of PSTV and CEV RNA labeled in vitro with ^{125}I has demonstrated that (i) each of these RNA species has a complexity compatible with the size estimate of 250-350 nucleotides and (ii) these RNA molecules do not have the same sequence. In the course of these studies we have developed a technique for the extraction of very small amounts (less than one microgram) of low molecular weight RNA from polyacrylamide gels of nucleic acid prep-

arations from healthy and viroid-infected plant tissues. The purity of such samples is high enough to allow subsequent iodination (to a specific activity of 1-5 x 10^6 cpm/μg) and characterization by two-dimensional fingerprinting analysis.

Our original interest in the study of viroids arose through a consideration of potential roles for RNA in the control of gene expression. Because viroids are composed entirely of RNA, it is possible that they could interact directly with the regulatory machinery of the host cells and could constitute aberrant forms of regulatory substances. In Chapter III we have considered the features of RNA which make it especially suitable for a regulatory role, then suggest a possible source for such proposed regulatory RNA, and finally describe possible modes of action. We have considered possible roles for extra RNA regions which are removed from RNA precursor molecules during their maturation, and have suggested that specific cleavage by RNA processing enzymes can give rise to reproducible fragments from these regions which may function after their removal. Nucleotide sequence analysis of an RNA processing site from bacteriophage T7 early messenger RNA precursor, as well as of potential cleavage sites from HeLa cell heterogeneous nuclear RNA, has been carried out.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

STUDIES OF IODINATED RNA

Studies on the in vitro iodination of a variety of nucleic acids with ^{125}I by Commerford (1971) demonstrated that such nucleic acids could be labeled to a high extent in a straightforward chemical reaction. Subsequently, Prensky et al. (1973) and Prensky (1975) described modifications of this procedure which allowed microgram quantities of RNA to be iodinated with carrier-free ^{125}I to specific activities in the range 10^6 to 10^8 counts per minute per microgram. With this development, it became possible to consider studying RNA species previously difficult or impossible to obtain labeled to high specific radio activity. Before such studies could begin, however, it was necessary to conduct a set of control experiments to study the properties of iodinated RNA molecules and to determine whether conventional techniques for the characterization of RNA could be applied successfully. This chapter, based partially upon a paper by H. D. Robertson, E. Dickson, P. Model and W. Prensky (1973), contains a description of these experiments.

The demands placed upon an in vitro RNA labeling method are, in general, quite stringent. Ideally the modified RNA should remain intact, retain its biological activity and preserve its chemical properties. In the studies to be reported here, we have shown that RNA fingerprinting and sequencing techniques originally developed using ^{32}P-labeled RNA species (Sanger et al., 1965; Brownlee et al., 1968; Brownlee and Sanger, 1969; Barrell, 1971) can be extended to include applications to RNA labeled in vitro with ^{125}I. Fingerprints of human 5S RNA and bacteriophage f2 RNA resemble those of their non-iodinated counterparts both in complexity and in specific pattern. The iodination method used shows little sensitivity to potential structure in single-stranded RNA molecules, yields stable oligonucleotide products in a reproducible manner, and does not change the specificity of

ribonucleases used conventionally for secondary sequence analysis. Thus in vitro iodination of trivial amounts (0.5 to 1.0 micrograms) of an unknown RNA species followed by fingerprinting will yield a unique pattern of spots (see page v for definitions of abbreviated terms commonly employed in fingerprinting analysis) within a two-dimensional fingerprint that would allow its identification at all future times. In addition, such fingerprints can be used for assaying the purity of RNA preparations, for comparisons with related RNA species, and as a starting point for sequence analysis. Some of these applications will be presented in Chapter II.

MATERIALS AND METHODS

Nucleic Acids for Iodination: RNA from bacteriophage f2 was grown and purified according to Webster et al. (1967). ^{14}C- and ^{32}P-labeled bacteriophage f1 messenger RNA was synthesized in vitro using the double-stranded Replicative Form DNA isolated from infected *E. coli* as template, according to the method of Model and Zinder (1974). Single-stranded DNA from phage f1 was the gift of K. Jakes, The Rockefeller University. We are indebted to Dr. T. Borun, Fells Research Institute, Philadelphia, Pa., for his gifts of purified HeLa cell 5S ribosomal RNA, and to Dr. W. Jelinek of The Rockefeller University, New York, for supplying us with in vivo ^{32}P-labeled HeLa cell 5S RNA. We received duck globin messenger RNA from Dr. D. Housman, Massachusetts Institute of Technology. Polyribocytidylic acid and polyuridylic acid were purchased from Sigma.

Isotopes: Carrier-free ^{125}I as NaI was purchased from New England Nuclear Corp. Carrier-free ^{32}P was obtained from Schwarz-Mann. Ribonucleoside triphosphate precursors for in vitro RNA synthesis ([^{14}C]CTP -- specific activity 500 mCi/mmole -- and alpha [^{32}P]ATP, GTP, UTP and CTP -- average specific activity 100,000 mCi/mmole) were obtained from New England Nuclear Corp.

Other materials: Ribonucleoside triphosphates were obtained from P-L Biochemicals Inc. BDH yeast nucleic acid for homochromatography was purchased from Gallard Schlesinger. RNases T1, T2 and U2 (Sankyo Co., Ltd., Tokyo, Japan) were purchased from CalBiochem. Pancreatic RNase (5x recrystallized), electrophoretically purified pancreatic DNase (DPFF), spleen phosphodiesterase, and snake venom phosphodiesterase (VPH) were obtained from Worthington Biochemical Corp.

Iodination of Nucleic Acids: All carrier-free iodinations with ^{125}I were carried out at the Sloan-Kettering Institute in collaboration with Dr. Wolf Prensky according to the methods of Commerford (1971), Prensky et al. (1973) and Prensky (1975). The concentrations and volumes of all reagents were the same as given by Prensky et al. (1973), except that some types of RNA were used in concentrations lower than 1 mg/ml. Also, all iodination products were separated not on hydroxyapatite but by use of CF11 cellulose (Franklin, 1966). The reaction mixture was brought to .5 ml with 0.1 M NaCl, 0.05 M Tris-HCl, pH 7.0, and 0.001 M ethylene diamine tetra-acetic acid; 0.5 ml of absolute ethanol was added, and the mixture loaded on CF11 cellulose, washed, and the [^{125}I]RNA eluted off in 0.9 ml of the same buffer less ethanol. The RNA was then ethanol precipitated by the addition of 10 micrograms of bacteriophage f2 RNA as carrier, 0.1 ml of a molar sodium acetate, pH 5.2 and 2.5 ml absolute ethanol. After incubation overnight at -20° C, centrifugation for 20 minutes at 15,000 rpm in a Sorvall SS 34 rotor yields visible pellets in siliconized 12 ml Sorvall pyrex tubes. The supernatants were poured off, the pellets dried in a vacuum dessicator to remove traces of ethanol, and the pellets resuspended in water for storage in siliconized screw-cap tubes at -20° C.

Digestion and Fingerprinting of Nucleic Acids: Iodinated RNA samples were digested and fingerprinted according to the standard techniques worked out by Sanger and his collaborators (Sanger et al., 1965; Brownlee and Sanger, 1969; Barrell, 1971). The ^{14}C-, ^{32}P- or ^{125}I-labeled RNA samples were mixed with 10 micrograms of bacteriophage f2 RNA and digested either with 2 micrograms of T1 ribonuclease in 2 micro-

liters of 0.01 M Tris-HCl (pH 7.5) - 0.001 M EDTA for 40 minutes at 37° C or digested with 2 micrograms pancreatic ribonuclease in the same buffer for 30 minutes at 37° C. Reaction mixtures were applied to cellulose acetate strips (Schleicher and Schuell, Keene, N.H.; 2.5 x 57 cm) soaked with 5% glacial acetic acid, containing 0.005 M EDTA in 7 M urea and subjected to high voltage electrophoresis (25 minutes, 6 kV).

The second dimensions involved either ascending homochromatography or high voltage electrophoresis. The oligonucleotides are transferred from the cellulose acetate to either Machery-Nagel 20 x 40 cm DEAE-cellulose thin-layer chromatography plates or to Whatman DE81 DEAE paper by placing the cellulose acetate strip in position on top of the DEAE substrate and applying wet Whatman 3MM strips with a strip of glass over the top to keep the wet layers flat. The water runs through the cellulose acetate into the dry DEAE layer carrying the oligonucleotides with it. The high affinity of nucleotides for DEAE causes them to bind tightly wherever they first contact the DEAE layer. Once transfer has been accomplished (about 20 minutes; three re-wettings of the Whatman 3MM strips), the glass, Whatman paper and cellulose acetate strip are removed carefully and the thin layer plate or DE81 paper rinsed in 95% ethanol for 5 minutes and air dried.

The thin layer plates were subjected to ascending homochromatography, a procedure carried out at 60° C in which a partially hydrolyzed solution of yeast RNA in 7 M urea moves upward over the DEAE-cellulose thin-layer plate causing the oligonucleotides to separate according to size, with the smallest ones moving the longest way up the plate. The original convention of placing the origin at the lower right corner in figures of homochromatographic fingerprints has been used (Brownlee and Sanger, 1969).

The DE81 DEAE paper was wetted with 7% formic acid for ^{32}P-labeled RNA or pH 1.9 buffer for ^{125}I-labeled RNA and subjected to high voltage electrophoresis. Dupont Cronex 2 X-ray film was used to detect the positions of all radioactive oligonucleotides in both types of fingerprints.

Oligonucleotides were eluted after two-dimensional fractionation according to conventional techniques (Barrell, 1971). Thirty percent triethylamine carbonate (TEC) in water (pH 8.7) was employed for elution of oligonucleotides from thin-layer plates, while a 22.5% solution was used to elute oligonucleotides from DEAE paper to obtain more rapid elution. Further base composition and enzymatic analyses were generally carried out exactly in accordance with the procedures reviewed by Barrell (1971), and any changes are noted in legends to figures describing individual experiments.

Homomix Preparation: Homomix is an abbreviated name for the homochromatography mixture described by Brownlee and Sanger (1969) for use in the second dimension of RNA fingerprints. It contains partially hydrolyzed nucleic acid dissolved in 7 M urea. The concentration of RNA and the extent of hydrolysis can be varied to produce a homomix of the appropriate strength (e.g. limited hydrolysis (0-5 minutes) yields a homomix capable of moving large oligonucleotides in a thin layer of DEAE-cellulose while extensive hydrolysis (10-25 minutes) results in a homomix that gives good resolution of the small oligonucleotides but leaves the larger ones at the origin). The most useful homomix for work presented here on RNA species ranging in size from 120 (HeLa 5S RNA) to 3500 (bacteriophage f2 RNA) was the "5%, 5 minute" homomix. 50 grams of BDH yeast nucleic acid (product no. 42045; from Gallard-Schlesinger) is added to 300 milliliters of freshly prepared 1 N KOH. The resulting sticky brown mass of crude RNA not yet dissolved in the KOH is stirred in a lively fashion with a heavy duty glass rod to expedite the dissolution of the RNA. A timer is set as soon as the RNA is added to the KOH. Exactly 5 minutes later (by this time the RNA is all in solution) the mixture is titrated to pH 7.5 with concentrated HCl. The neutralized mixture is transferred to two three-foot lengths of 1¼-inch dialysis tubing (washed briefly in distilled water) leaving a space of 3-4 inches of empty tubing between the top of the liquid and the top knot of the dialysis tubing. This is dialysed against 4 liters of water for two hours. The water is replaced with fresh

incubation is for one hour. The samples are then spotted and run in the same way as the products of alkaline hydrolysis at pH 3.5 on Whatman 3MM paper.

RESULTS

Fingerprinting of Iodinated RNA

Two-dimensional fractionation of the specific oligonucleotides resulting from digestion of an RNA with T1 or pancreatic RNase yields a unique two-dimensional pattern (called a fingerprint) for each different RNA species analyzed. The fact that the pattern obtained is extremely sensitive to single base changes in the RNA allowed Robertson and Jeppesen (1972) to determine the extent of variation in three closely related bacteriophage RNA molecules. In these studies the RNA genomes of bacteriophage f2, MS2 and R17 were shown to differ from each other by about 4%. Although these RNA species shared 96% of their sequences, two-dimensional fingerprints of their ribonuclease T1-resistant oligonucleotides exhibited totally different patterns in the characteristic region where large fragments are situated.

We have tested iodinated RNA to see whether it is compatible with fingerprinting procedures conventionally in use with ^{32}P-labeled molecules such as those studied by Robertson and Jeppesen (1972).

Figure 1 presents a comparison of fingerprints obtained following ribonuclease T1 digestion of either ^{32}P- or ^{125}I-labeled bacteriophage f2 RNA. Figure 1A shows the well characterized fingerprint of ^{32}P-labeled f2 RNA isolated from in vivo ^{32}P-labeled phage particles (Robertson and Jeppesen, 1972). The bracketed spots represent a unique group of T1-resistant oligonucleotides which occur only once each in the RNA molecule. Figure 1B shows that f2 RNA labeled with ^{125}I yields a very similar fingerprint to that in Figure 1A and that, in particular, the unique spots are arrayed in a similar pattern. Thus, for an RNA species such as f2 with a complexity of 3500 nucleotides, the iodinated form can be digested by T1 ribonuclease to yield a specific

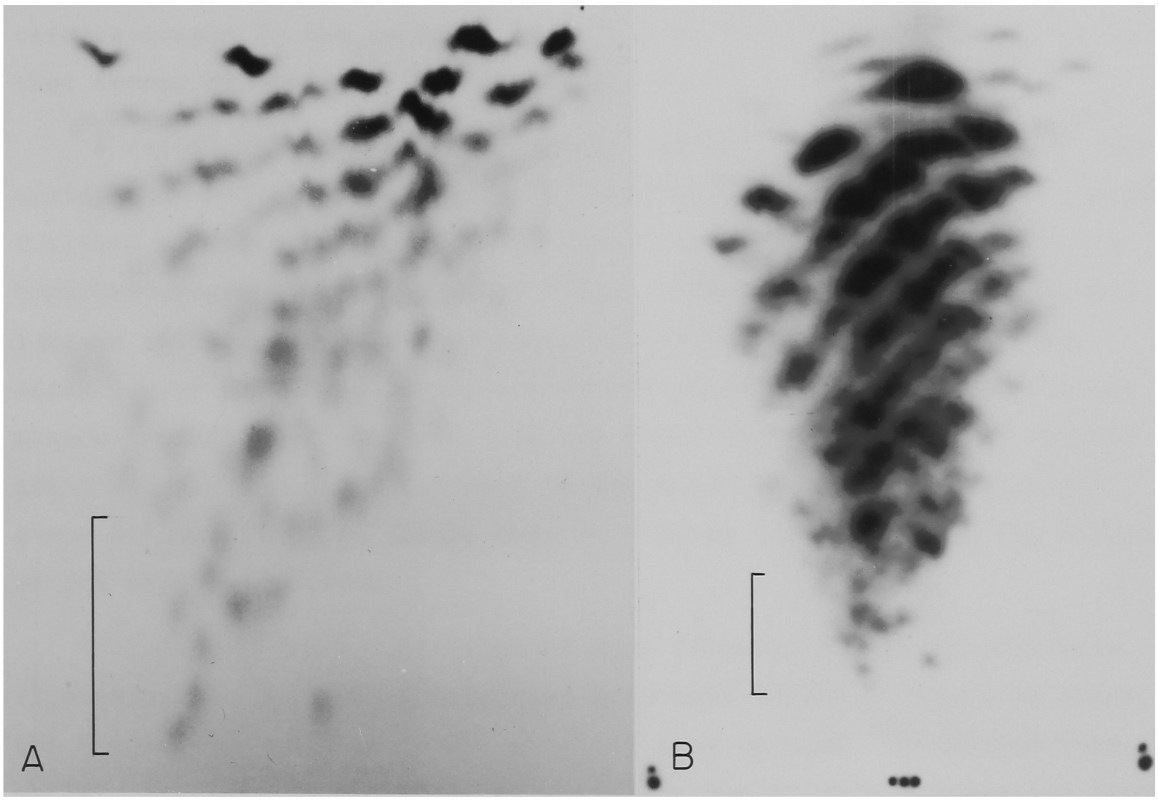


Figure 1

Ribonuclease T1 fingerprints of ^{32}P- and ^{125}I-labeled f2 phage RNA. Approximately 5 x 10^6 dpm of either ^{32}P-labeled f2 phage RNA (specific activity 2 x 10^6 dpm/microgram) or ^{125}I-labeled f2 RNA (specific activity between 0.3-1.1 x 10^7 dpm/microgram) prepared as described in Materials and Methods were mixed with 10 micrograms of unlabeled f2 phage RNA and digested with 2 microliters of T1 ribonuclease for 40 minutes at 37° C. Fingerprinting was then carried out employing thin-layer homochromatography (see Materials and Methods). (A) RNase T1 fingerprint of ^{32}P-labeled f2 RNA. (B) RNase T1 fingerprint of ^{125}I-labeled f2 RNA. In each case the origin is at the lower right, and the bracketed region near the bottom of each picture represents that population of unique oligonucleotides in each fingerprint which correspond to those characterized previously (Robertson and Jeppesen, 1972). The different size of the two brackets is explained by the somewhat different spread of comparably sized oligonucleotides obtained in the two fingerprints which were run at different times.

population of resistant oligonucleotides which after electrophoretic separation in the first dimension and separation by ascending homochromatography in the second dimension form a pattern quite similar to that formed by their ^{32}P-labeled counterparts.

A second method of RNA fingerprinting employs electrophoretic separations in both dimensions (pH 3.5 electrophoresis on cellulose acetate followed by electrophoresis in 7% formic acid on DEAE paper). Unlike homochromatography, this type of fingerprint can resolve many sequence isomers and has highly reproducible mobilities of oligonucleotides which depend on the base composition and thus provide important information for sequence analysis. On the other hand, this type of fingerprint is only useful in the case of RNA molecules of low complexity since oligonucleotides longer than 8-10 bases long remain at the origin of the second dimension.

Figure 2 shows a comparison of such fingerprints of ribonuclease T1 digests of ^{32}P- and ^{125}I-labeled HeLa cell 5S RNA. Figure 2A shows the fingerprint of HeLa 5S RNA obtained from in vivo ^{32}P-labeled HeLa cells. The pattern formed by oligonucleotides in this fingerprint bears a striking resemblance to the fingerprint of ^{32}P-labeled 5S RNA isolated from KB cells (Forget and Weissman, 1967). Figure 2B shows the pattern obtained when HeLa cell 5S RNA is iodinated and subjected to similar treatment. In this case, however, the second dimension was run at pH 1.9 with .005 molar EDTA instead of in 7% formic acid (pH 1.7) since the iodinated oligonucleotides tend to streak in a "7%" second dimension. While certain similarities in pattern can be picked out in these two fingerprints, it does not appear to be an ideal system for comparison of RNA species since the most characteristic oligonucleotides (i.e., the longest T1-resistnat oligonucleotides) are clustered either on the origin or close to it.

T1 fingerprints of HeLa 5S RNA employing homochromatography as the second dimension are shown in Figure 3. These two patterns are identical except that 4 spots present in the fingerprint of ^{32}P-labeled RNA are absent from the fingerprint of ^{125}I-labeled RNA.

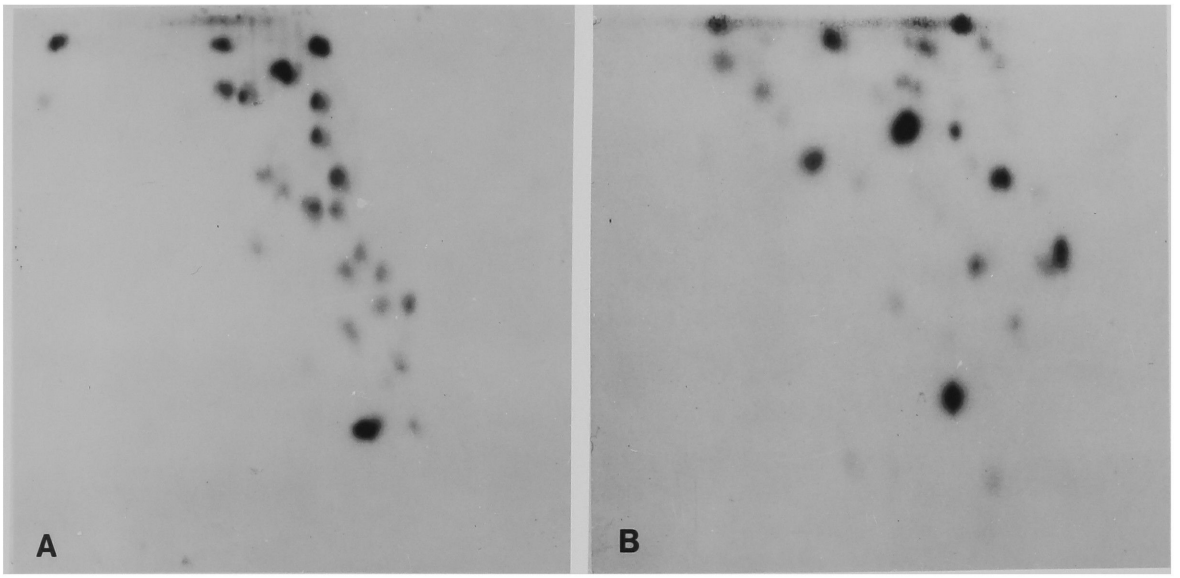


Figure 2

Ribonuclease T1 fingerprints of ^{32}P- and ^{125}I-labeled HeLa cell 5S RNA with electrophoretic second dimension. Approximately 2 x 10^6 dpm of ^{32}P- or ^{125}I-labeled HeLa cell 5S RNA (specific activity 0.5 x 10^6 and 1.1 x 10^7 dpm/microgram respectively) were mixed with 10 micrograms f2 phage RNA and digested with RNase T1. Two-dimensional fingerprints in which the first dimension consisted of electrophoresis at pH 3.5 on cellulose acetate strips and the second dimension was electrophoresis at about pH 1.7 in 7% formic acid for the ^{32}P-labeled 5S RNA or in acetic-formic buffer, pH 1.9, for the ^{125}I-labeled 5S RNA (see Material and Methods for details). (A) Two-dimensional fingerprint of ^{32}P-labeled HeLa 5S RNA: the origin is at the upper right in this representation; the first dimension was from right to left, and the second (high voltage electrophoresis in 7% formic acid on Whatman DE81 paper) from the top of the picture to the bottom. (B) Two-dimensional fingerprint of ^{125}I-labeled HeLa 5S RNA carried out exactly as the fingerprint in (A) except that the second dimension was run at pH 1.9 in acetic-formic acid buffer, pH 1.9 - 0.005 M EDTA.

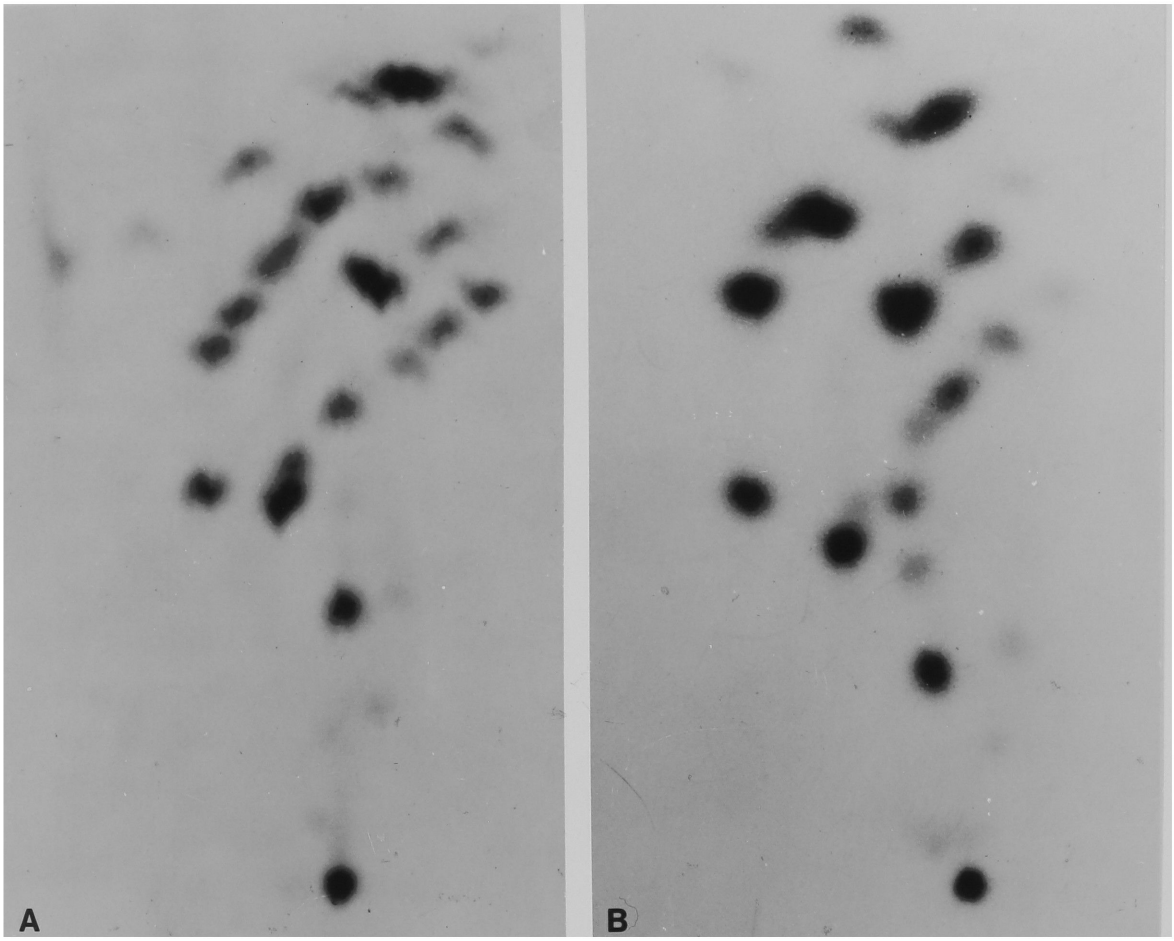


Figure 3

Two-dimensional fingerprints of RNase T1-resistant oligonucleotides of ^{32}P- and ^{125}I-labeled HeLa 5S RNA. RNA samples containing approximately 2 x 10^6 dpm of ^{32}P- or ^{125}I-labeled HeLa 5S RNA (specific activity 0.5 x 10^6 and 1.1 x 10^7 dpm/microgram respectively) were digested with RNase T1 and subjected to fingerprinting with a thin-layer homochromatographic second dimension as described in Materials and Methods. (A) Fingerprint of ^{32}P-labeled HeLa 5S RNA. (B) Fingerprint of ^{125}I-labeled HeLa 5S RNA. The origin is at the lower right; the first dimension was from right to left, and the second, from the bottom to the top.

Thus we find that iodinated RNA can be subjected to all of the procedures required for fingerprinting by either of the two methods in general use, and furthermore, that fingerprints of ^{32}P- and ^{125}I-labeled RNA can be compared directly.

Base Specificity of Iodination

Commerford (1971) has shown that the most efficiently labeled product of iodination of homopolymers is 5-iodocytosine while uracil and guanine are labeled at 3-4% and 0.2-0.3% of the level of cytosine respectively. Poly(C) and poly(U) were iodinated, reduced to monomers by enzymatic digestion with RNase T2 (containing RNase T1 and pancreatic RNase), and characterized by high voltage electrophoresis at pH 3.5 on Whatman 3MM paper (Figure 4A, lanes 3 and 5). Under these conditions, ^{125}I-labeled CMP runs slightly slower than ^{32}P-labeled GMP (Figure 4A, lane 1), while ^{125}I-labeled UMP has a mobility midway between ^{32}P-labeled GMP and ^{32}P-labeled UMP. Similar digestion of iodinated HeLa 5S RNA yields a major component with mobility identical to that of ^{125}I-labeled CMP, while a trace spot is observed at the mobility of ^{125}I-labeled UMP (Figure 4B, lane 1). The same result was obtained in base composition determinations of RNA's from diverse sources. Transfer RNA provides an exception to this observation. In this case T2 RNase digestion yields a complicated pattern of up to nine ^{125}I-labeled components (Figure 4C). It is possible that in vitro labeling with ^{125}I will prove to be a valuable technique for identifying modified bases in such species.

A second method commonly used for the determination of base composition of RNA is alkaline hydrolysis (incubation in 0.4 N NaOH, at 37° for 18 hours). This treatment has been found unsuitable for iodinated species for two reasons. First, a significant portion of the ^{125}I is lost from ^{125}I-labeled poly(C) during this treatment. Second, upon electrophoresis of iodinated RNA's following alkali treatment, there is a shift of most of the radioactivity from the position of ^{125}I-labeled CMP to that of ^{125}I-labeled UMP (Figure 4A, lane 2). Similar treatment

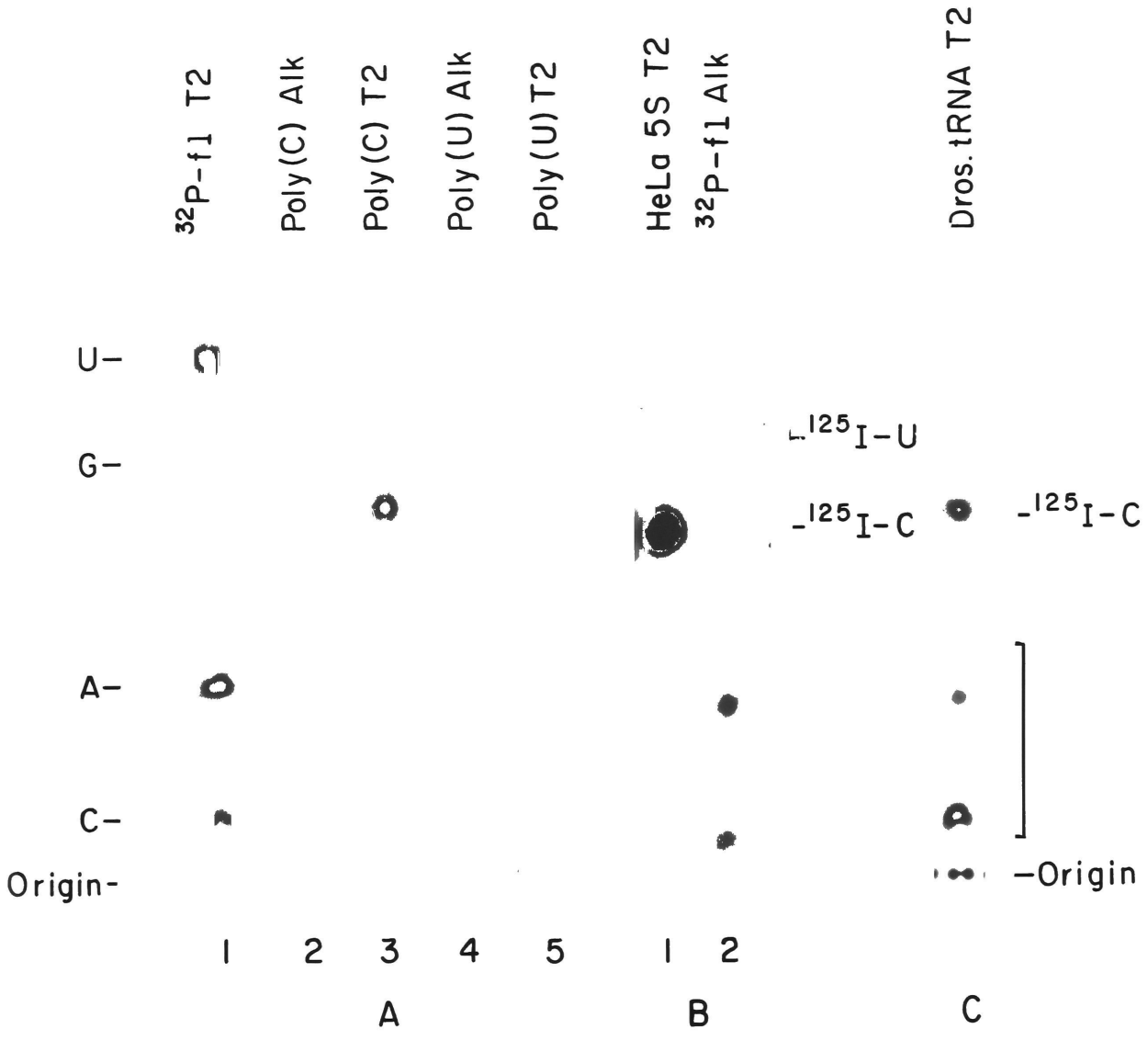
Figure 4

Base composition analysis of iodinated nucleic acid species. Alkaline hydrolysis was for 18 hours at 37° in 0.4 M NaOH, in a volume of ten microliters and this treatment is abbreviated as "Alk" in the figure. RNase T2 was employed at a concentration of two units/ml together with RNase T1 (0.05 mg/ml) and pancreatic RNase (0.05 mg/ml). Incubation was for 2 hours at 37° in ten microliters of buffer containing 0.05 M ammonium acetate, pH 4.5. This treatment is abbreviated as "T2" in the figure. Various samples were spotted onto Whatman 3MM paper and exposed to descending high-voltage electrophoresis at pH 3.5 in 0.5% pyridine, 5% glacial acetic acid, 0.005 M EDTA.

(A) lane 1--RNase T2 digestion of ^{32}P-labeled f1 RNA (10^4 dpm; specific activity 2 x 10^8 dpm/microgram) prepared with all four α-^{32}P-labeled ribonucleoside triphosphates; lane 2--alkaline hydrolysis of ^{125}I-labeled poly(C) (10^4 dpm; specific activity 2.5 x 10^6 dpm/microgram); lane 3--RNase T2 digestion of ^{125}I-labeled poly(C) (10^4 cpm as in lane 2); lane 4--alkaline hydrolysis of ^{125}I-labeled poly(U) (5 x 10^3 dpm/ specific activity 10^5 dpm/microgram); lane 5--RNase T2 digestion of ^{125}I-labeled poly(U) (5 x 10^3 dpm as in lane 4).

(B) lane 1--RNase T2 digestion of ^{125}I-labeled HeLa 5S RNA (10^4 dpm; specific activity 1.1 x 10^7 dpm/microgram); lane 2--alkaline hydrolysis of ^{32}P-labeled f1 RNA (10^4 dpm as in (A), lane 1).

(C) RNase T2 digestion of 2 x 10^4 dpm of ^{125}I-labeled Drosophila tRNA (specific activity greater than 10^6 dpm/microgram). The bracketed spots represent additional ^{125}I-labeled moieties which we have observed only in RNase T2 digests of a number of iodinated RNA samples which contained tRNA.


^{32}P-f1 T2
Poly(C) Alk
Poly(C) T2
Poly(U) Alk
Poly(U) T2
HeLa 5S T2
^{32}P-f1 Alk
Dros. tRNA T2
U–
G–
A–
C–
Origin–
^{125}I-U
^{125}I-C
^{125}I-C
–Origin
1 2 3 4 5 1 2
A
B
C

of ^{125}I-labeled poly(U) results in a slight loss of radioactivity, but no change in mobility of the resulting ^{125}I-labeled UMP (Figure 4A, lane 4). These observations could be explained if either one or the other of two possible reactions--either deamination or de-iodination of ^{125}I-labeled CMP--occurs under extreme alkaline conditions.

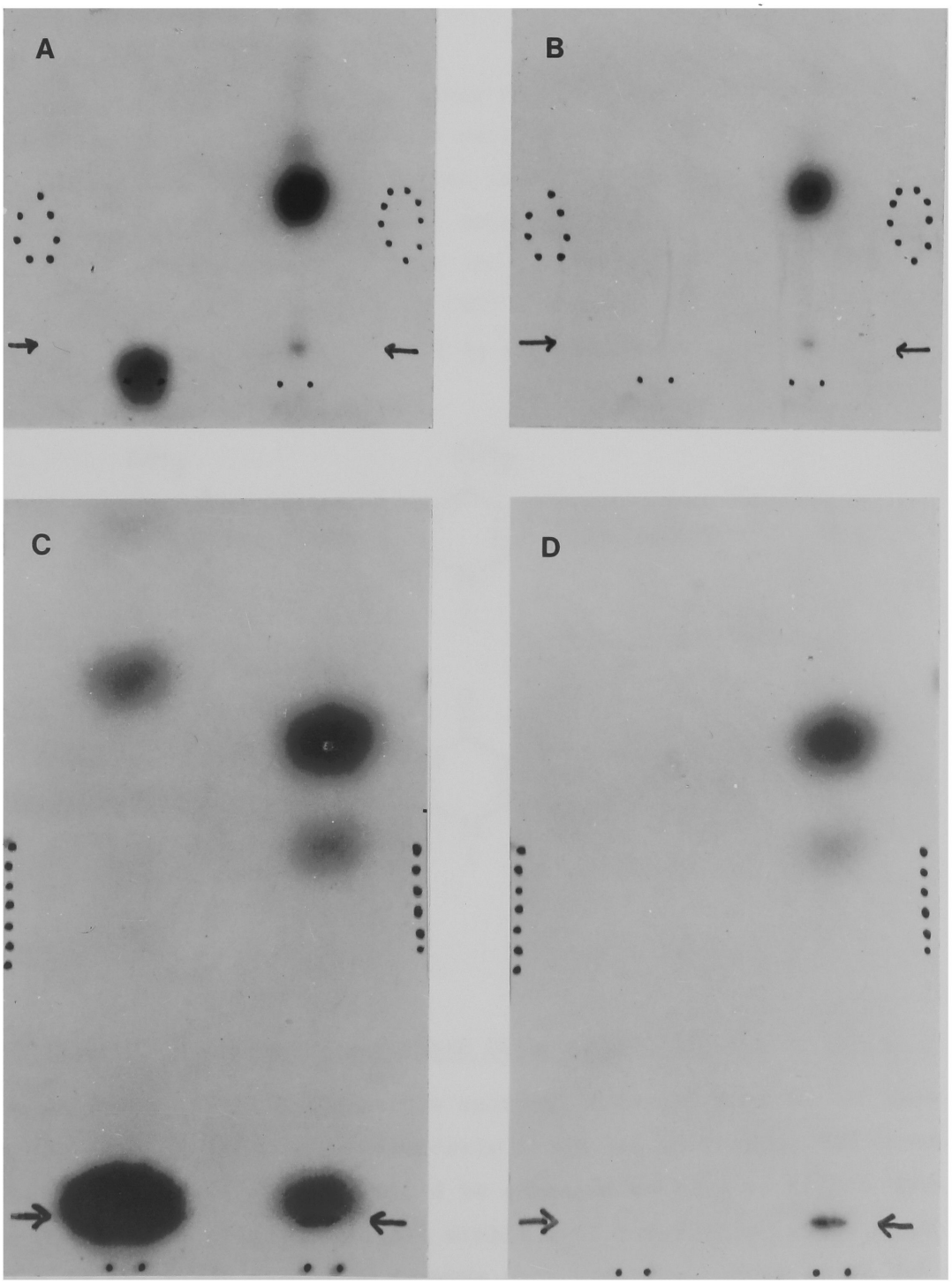
An experiment to test the possibility that these chemical changes do occur in iodinated cytidylate residues is presented in Figure 5. Bacteriophage f1 RNA containing ^{14}C-labeled CMP, synthesized in vitro with RNA polymerase from a double-stranded Replicative Form DNA template, was divided into two aliquots; half was iodinated. Both samples were digested with RNase T2, spotted on Whatman 3MM paper and subjected to high voltage electrophoresis at pH 3.5 as described in Materials and Methods. Figure 5A represents the X-ray film which was in contact with the 3MM paper. Lane 1 shows the base composition of the ^{14}C-labeled RNA (non-iodinated) while lane 2 shows the base composition of the ^{125}I-labeled aliquot. As expected, the non-iodinated form of CMP runs near the origin and the iodinated form runs slightly faster than the blue dye marker (xylene cyanol FF) whose position is indicated by the circle of dots at each side. Figure 5B represents a second piece of X-ray film that was exposed to the 3MM paper but separated from it by the thickness of the first film. Thus, the ^{14}C is not detected and only the ^{125}I-labeled CMP can be seen. Both the iodinated and non-iodinated spots of CMP were eluted from the paper, subjected to alkaline hydrolysis (0.4 M NaOH, 18 hours, 37° C), respotted onto 3MM paper, and re-electrophoresed at pH 3.5. Again two films were exposed simultaneously. The closest film to the paper (Figure 5C) shows that alkaline hydrolysis of iodinated CMP (lane 2) has caused a conversion of almost all of the radioactivity that had previously migrated at the position of ^{125}I-labeled CMP to two new products: one which migrates in the position of non-iodinated CMP near the origin, the other which migrates in the position of ^{125}I-labeled UMP (see Figure 4A, lanes 4 and 5). Furthermore, the radioactive oligonucleotide migrating near the origin has not been able to expose the second film

Figure 5

Deamination and de-iodination of iodo CMP under alkaline conditions. Bacteriophage f1 RNA containing ^{14}C-labeled CMP was digested to mononucleotides by incubation with RNase T2, spotted onto Whatman 3MM paper, and subjected to high voltage electrophoresis at pH 3.5. Its location near the origin was detected by exposure to X-ray film (panel A, lane 1). A second aliquot of this RNA which had first been iodinated in vitro with ^{125}I was treated identically (panel A, lane 2). In this case, the base composition indicates the presence of iodo-CMP with a trace of radioactivity in the position of iodo-UMP (see Figure 4B, lane 1). (B) represents the second film exposed to this same 3MM paper but separated from the paper by the thickness of one X-ray film (that shown in (A)). In this case, spots containing only ^{14}C are not visible. Only the spots in lane 2 which contain ^{125}I can be seen.

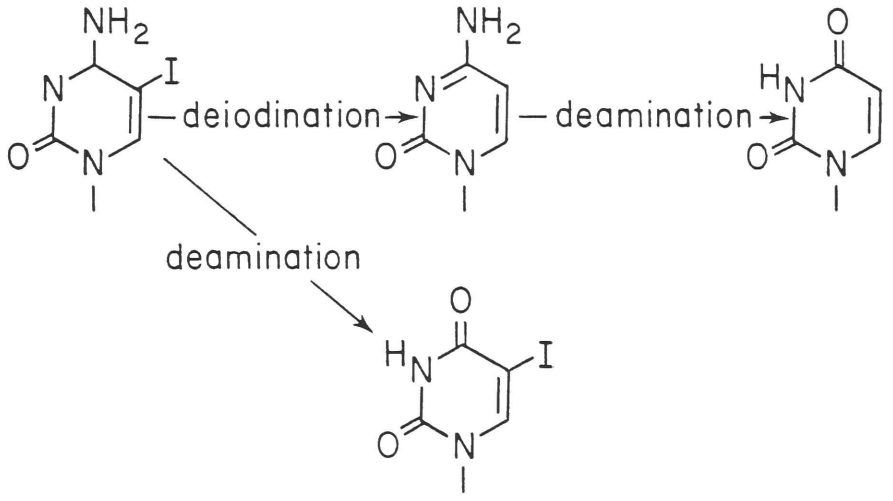
The two major spots shown in (A) were eluted, and subjected to alkaline hydrolysis and electrophoresis as above. (C) represents the X-ray film that was directly in contact with the 3MM paper and (D) represents the second film. The ^{14}C-labeled CMP in lane 1 near the origin has lost a small minority of its radioactivity to a second spot that runs in the position of UMP (see Figure 4A, lane 1). Neither of these spots are visible on the second film. On the other hand, the iodo-CMP shown in (C), lane 2, has converted most of its radioactivity into two new species: one that migrates just like iodo-UMP (see Figure 4A), and the other that runs exactly with the non-iodinated CMP shown in lane 1. Comparison of (C) with (D) shows that one of these new spots (the one migrating in the position of iodo-UMP) contains ^{125}I. The other new spot which co-migrates with CMP contains only ^{14}C.

Details of digestion and electrophoresis of the RNA are described in Materials and Methods.


A
B
C
D

(Figure 5D) indicating that the radioactivity present is ^{14}C only. This result confirms the proposed conversion under alkaline conditions of 5-iodo-CMP to either 5-iodo-UMP (by deamination) or to de-iodinated CMP (by de-iodination).

Lane 1 of Figure 5C demonstrates the fact that this same reaction takes place, but at a much reduced rate, in the case of non-iodinated CMP. Here, some ^{14}C can be detected in the position of UMP (see Figure 4A, lane 1), while the vast majority remains near the origin. Apparently the presence of ^{125}I at the C_5 position of the pyrimidine ring (ortho to the amino group of CMP) increases the probability of nucleophilic attack at both C_4 and C_5 under alkaline conditions.



Mobility of Iodinated Oligonucleotides

As discussed in the previous section, iodo-CMP runs faster than CMP when subjected to electrophoresis at pH 3.5 on Whatman 3MM paper. This difference in mobility would be expected to have an effect upon the position of oligonucleotides within a T1 fingerprint since separation in the first dimension depends upon electrophoresis at pH 3.5 on

cellulose acetate. To investigate the impact of such a mobility shift upon the final form of a fingerprint, bacteriophage f1 RNA which had been transcribed in vitro in the presence of ^{14}C-labeled CMP was divided into two aliquots, half was iodinated with ^{125}I, and both halves were subjected to T1 fingerprinting analysis as shown in Figure 6. Panel (A) shows the fingerprint of ^{14}C RNA alone. Panels (B) and (C) represent the film in contact and the second film exposed to the fingerprint of the ^{125}I-^{14}C-labeled RNA mixed with some non-iodinated ^{14}C-labeled RNA. As predicted, the iodinated oligonucleotides move faster in the first dimension (from right to left) causing the fingerprint in (C) to be shifted further to the left with respect to the blue dye marker (shown as a circle of dots in the upper right corner) than the fingerprint in (A). Panel B shows these two effects superimposed upon each other. This demonstrates that the shift is most dramatic with the shortest oligonucleotides (near the top of the fingerprint), and becomes increasingly less apparent in the larger ones. Furthermore, it is evident that the acquisition of a single ^{125}I is insufficient to alter significantly the mobility of the oligonucleotides in the second homochromatographic dimension. In order to find out whether the iodinated oligonucleotides retain the same spatial arrangement within the fingerprint as that demonstrated by non-iodinated oligonucleotides, it is necessary to carry out secondary analysis of all the spots from a fingerprint of an RNA species whose sequence is known.

Secondary Analysis of Iodinated Oligonucleotides

Because we had a source of both unlabeled and ^{32}P-labeled HeLa 5S RNA and because the nucleotide sequence of human 5S RNA from KB cells is known (Forget and Weissman, 1967), we chose this RNA species to test the conventional secondary analysis procedures and to determine whether fingerprints of ^{32}P- and ^{125}I-labeled RNA species can be compared directly, spot for spot.

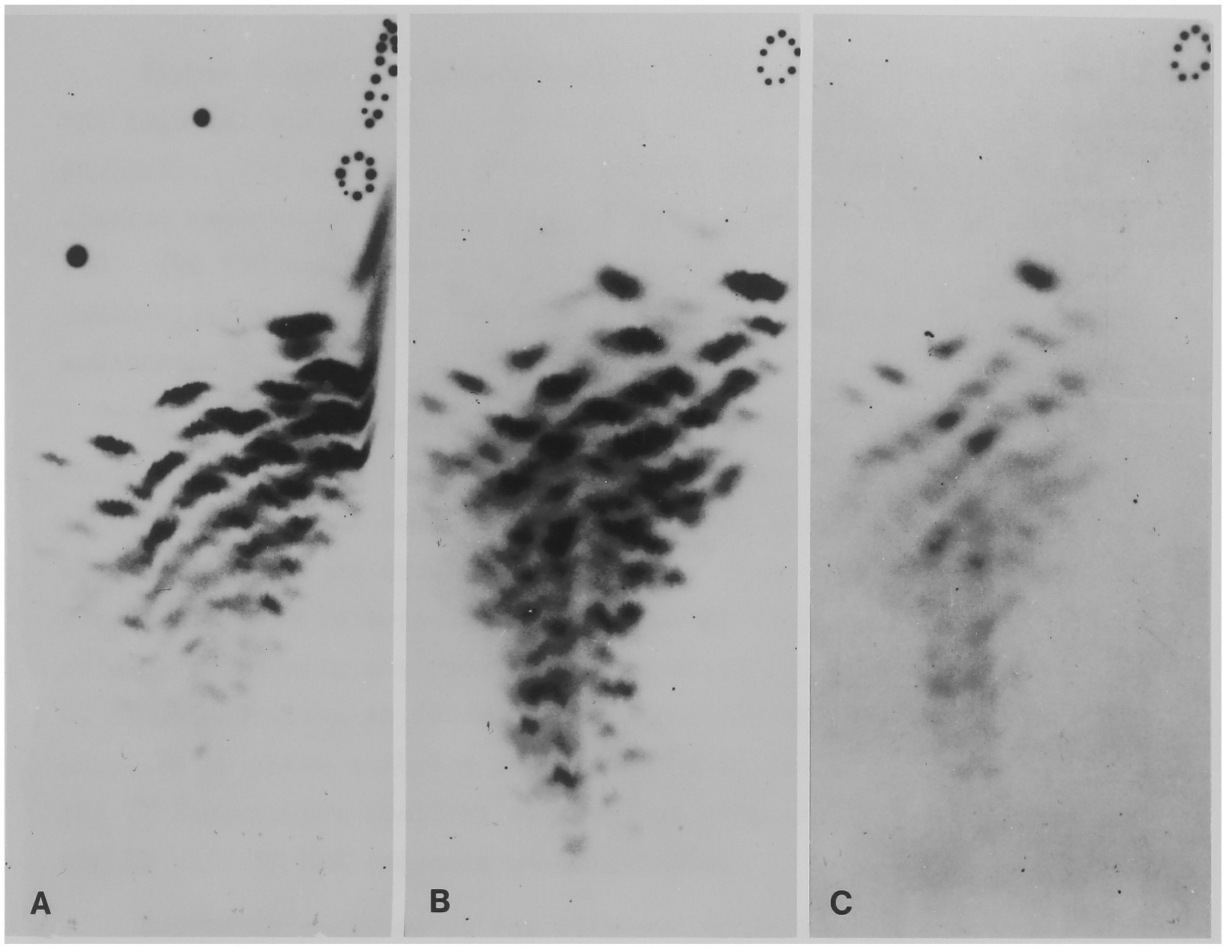


Figure 6

Mobility shift of iodinated oligonucleotides. A sample of bacteriophage f1 RNA (specific activity 5 x 10^4 dpm/microgram synthesized in vitro with ^{14}C-labeled CTP) was iodinated with ^{125}I (^{125}I specific activity about 5 x 10^7 cpm/microgram). Aliquots of both non-iodinated and iodinated forms were subjected to T1 fingerprinting analysis. The iodinated RNA sample was mixed with an aliquot of the non-iodinated ^{14}C(CMP)-labeled RNA. The positions of both ^{14}C-only and ^{125}I-containing oligonucleotides can be seen in (B) which represents the X-ray film in contact with this fingerprint. A whole set of rightward spots visible in (B) can not be seen in (C) (the second film on this fingerprint) thus allowing us to attribute their presence to ^{14}C-labeled oligonucleotides which are not iodinated. The circle of dots in the upper right hand corner marks the position of the blue dye marker. Panel (A) shows a T1 fingerprint of ^{14}C-labeled f1 RNA which had not been subjected to iodination for comparison.

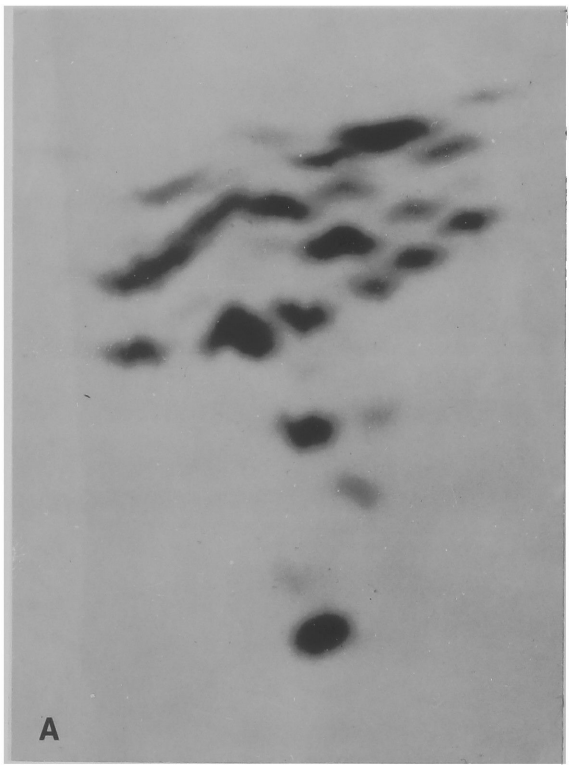
Figure 7 shows T1 fingerprints of ^{32}P- and ^{125}I-labeled HeLa 5S RNA together with drawings indicating the spots selected for secondary analysis. The surface of the thin layer plate within each of the circled regions was scraped off and the RNA eluted from it with 30% TEC. The TEC was removed by vacuum dessication and the retrieved radioactivity was split into the appropriate number of aliquots for subsequent analysis.

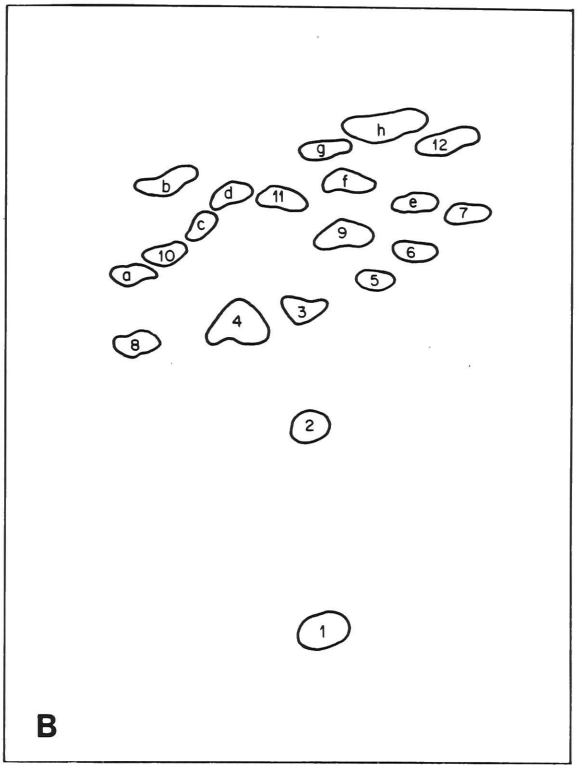
It was first necessary to assign sequences to all of the oligonucleotides present in the fingerprint of the ^{32}P-labeled HeLa 5S RNA to see if it is the same as the 5S isolated from human KB cells and sequenced by Forget and Weissman (1967). The oligonucleotides indicated in Figure 7B were eluted from the T1 fingerprint and subjected to electrophoretic analysis following RNase T2, pancreatic RNase, or U2 RNase digestion as described in Materials and Methods. The data obtained in these analyses are presented in Table I. Each spot from the T1 fingerprint could be assigned an oligonucleotide sequence from the KB cell 5S RNA sequence unambiguously.

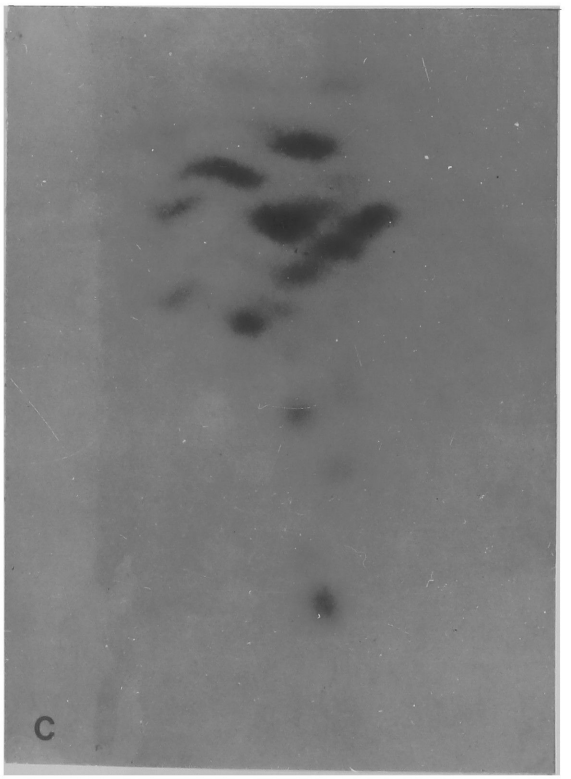
Enzymatic analyses of the oligonucleotides extracted from the regions circled in the tracing of the ^{125}I T1 fingerprint of HeLa 5S RNA (Figure 7D) are presented in Figure 8. Each eluted oligonucleotide sample was divided into four equal aliquots. One aliquot was left undigested and the other three were digested with RNase U2, pancreatic RNase or T2 RNase as described in the Materials and Methods. These four samples were then spotted side by side on DE81 DEAE-paper and subjected to high voltage electrophoresis at pH 1.9. By combining information from position in the fingerprint, mobility when rerun at pH 1.9 (with no prior digestion), and number and mobility of digestion products of U2 and pancreatic RNases at pH 1.9, with knowledge of the sequences of C-containing oligonucleotides present in 5S RNA, it was possible to assign a sequence to each of the oligonucleotides eluted with one exception (to be dealt with below). These data and assignments are presented in Table II. Each oligonucleotide whose sequence could be identified occupied the same position within

Figure 7

T1 fingerprints of ^{32}P- and ^{125}I-labeled HeLa 5S RNA subsequently subjected to secondary enzymatic analysis. (A) T1 fingerprint of ^{32}P-labeled HeLa 5S RNA obtained from Dr. W. Jelinek of The Rockefeller University. (B) Drawing indicating the spots within the fingerprint shown in (A) which were selected for secondary enzymatic analysis. (C) T1 fingerprint of ^{125}I-labeled HeLa 5S RNA (unlabeled RNA was obtained from Dr. T. Borun, Wistar Inst., Philadelphia). (D) Drawing indicating the spots within the fingerprint shown in (B) which were selected for secondary enzymatic analysis.


A


h
12
g
f
b
d
11
e
7
c
9
10
6
a
5
3
4
8
2
1
B


C

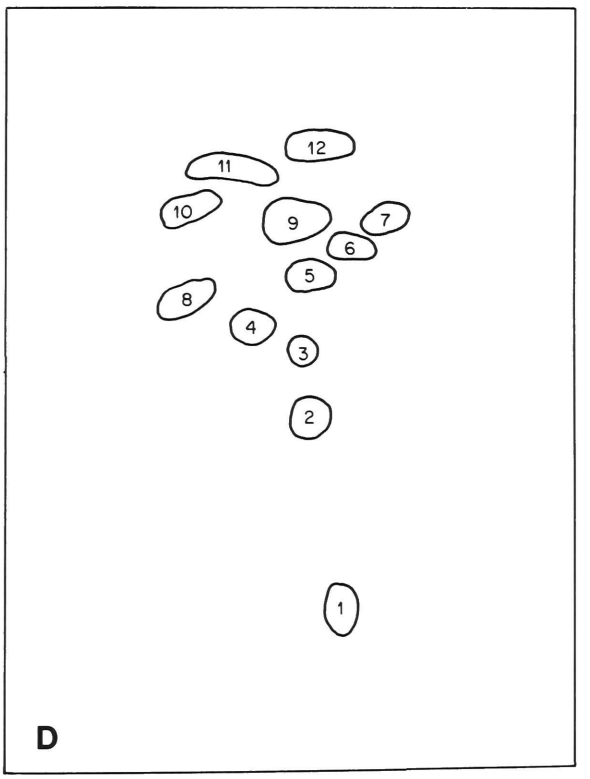
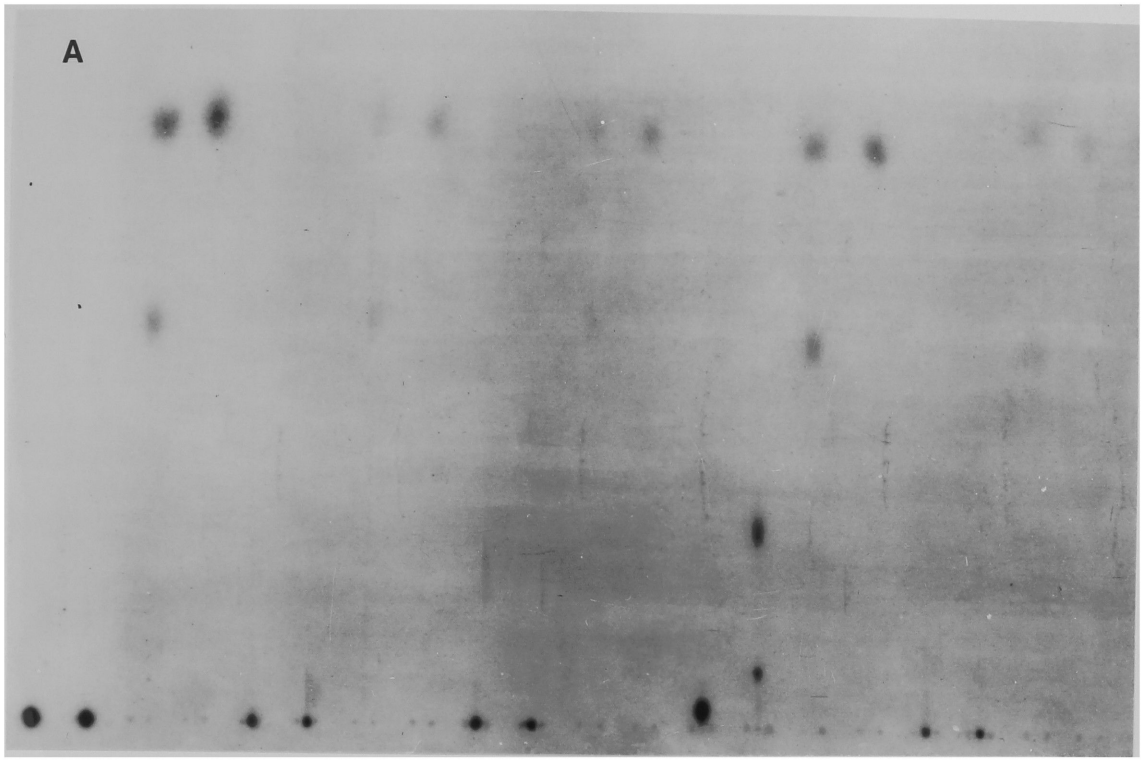
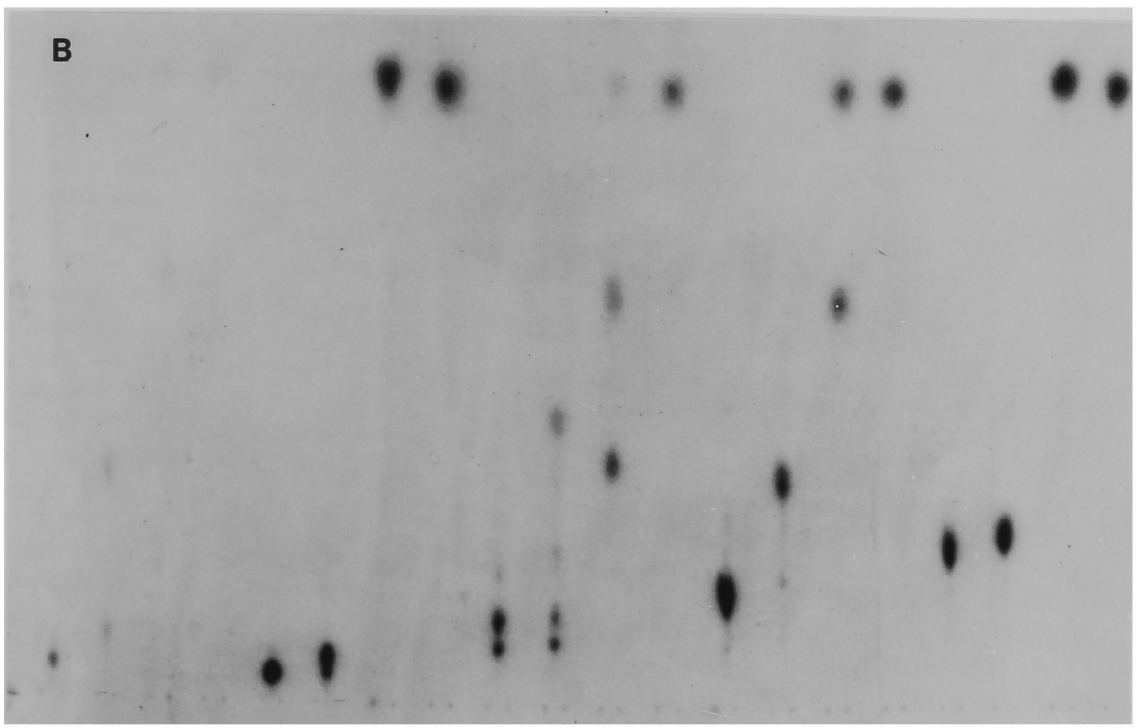

12
11
10
9
7
6
5
8
4
3
2
1
D

Table I

Sequence assignment of RNase T1-resistant oligonucleotides from ^{32}P-labeled HeLa 5S RNA (see Figures 3A and 7A)

Oligonucleotide number	Proposed Sequence (Forget & Weisman, 1967)	Confirmatory Secondary Analyses: Base Composition (RNase T2)	Pancreatic RNase Digestion Products	RNase U2 Digestion Products
1	CCAUACCACCCUG	C,A,G,U	AU,AC,C,G,U	CCCUG,UA,CCA
2	AAUACCG	C,A,G,U	AAU,AC,C,G	CCG,UA,A
3	CUAAG	C,A,G,U	AAG,C,U	CUA,G,A
4	AUCUCG) mixture UCUACG)	C,A,G,U	AU,AC,G,C,U	UCUCG,UCUA, UA,CG,A
5	AACG	C,A,G	AAC,G	CG,A
6	ACCG	C,A,G	AC,G,C	CCG,A
7	CCCG	C,G	G,C	CCCG
8	UACUUG	C,A,G,U	AC,G,U	CUUG,UA
9	CCUG) mixture AAG)	C,G,U A,G	G,C,U AAG	CCUG A,G
10	UCUG	C,G,U	G,C,U	UCUG
11	CUG) mixture UCG)	C,G,U	G,C,U	CUG UCG
12	CG	C,G	G,C	CG
a	UUAG	A,G,U	AG,U	UUA,G
b	UG	G,U	G,U	UG
c	AUG) mixture CUG/UCG)	C,A,G,U	AU,G,C,U	UG,A CUG/UCG
d	UAG	A,G,U	AG,U	UA,G
e	CAG	C,A,G	AG,C	CA,G
f	AG	A,G	AG	G,A
g	CUU_{OH}	C,U	C,U	CUU_{OH}


A


B

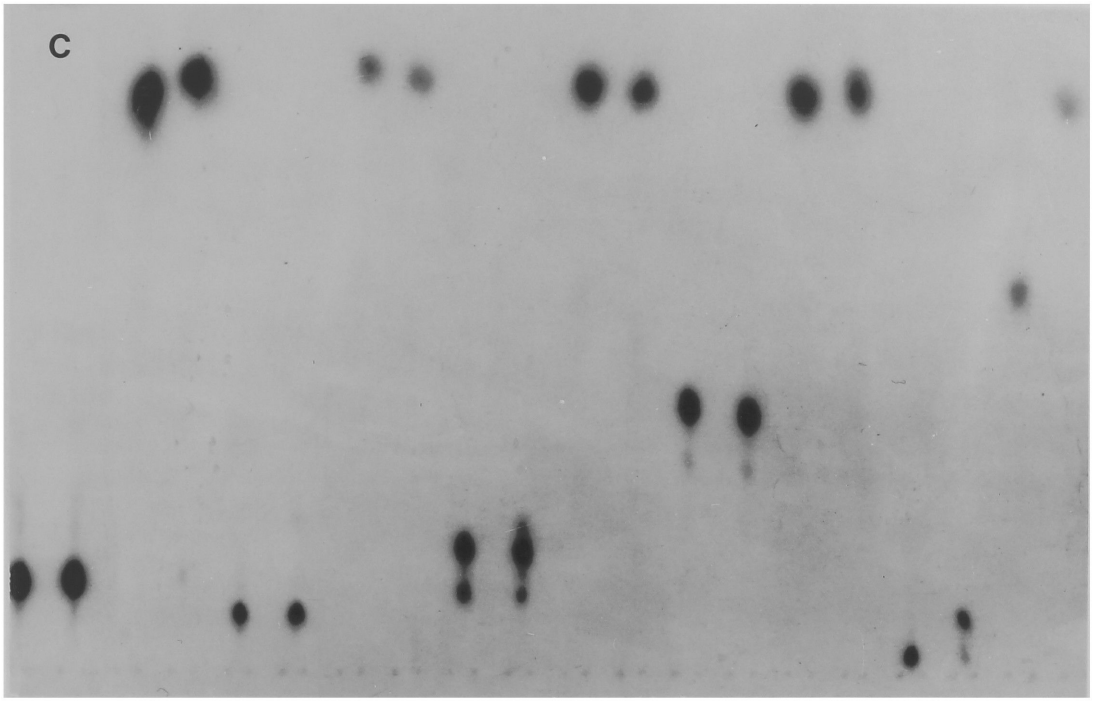


Figure 8

Enzymatic analysis of RNase T1-resistant oligonucleotides from ^{125}I-labeled HeLa 5S RNA. Oligonucleotides circled in Figure 7D were eluted, split into four equal aliquots, and treated as follows: lane 1--no treatment; lane 2--incubation with RNase U2 in ten microliters at a concentration of 2 units/ml for 2 hours at 37° C in 0.05 M sodium acetate pH 4.5, 0.002 M EDTA, 0.1 mg/ml crystallized bovine serum albumin (Pentex Inc., Kankakee, Ill.); lane 3--incubation with pancreatic RNase (1 mg/ml) in ten microliters of 0.01 M Tris-HCl, pH 7.4, 0.001 M EDTA; lane 4--treatment with RNase T2 in the enzymatic mixture described in the Materials and Methods. The fifteen groups of four analyses are labeled with the numbers 1 to 12 to correspond to oligonucleotides indicated in Figure 7B, and x, y, z to designate additional minor oligonucleotides analyzed from the fingerprint in Figure 3B (data not presented). The analyses of oligonucleotides numbered in Figure 7B are distributed among the three parts of this figure in the following way: (A) 1, x, y, 2, z; (B) 3, 4, 5, 6, 7; (C) 9, 10, 11, 12, 8 (reading from left to right across the origin). The reactions were spotted onto Whatman DE81 DEAE-paper and subjected to high-voltage electrophoresis in pH 1.9 buffer (2.5% formic acid, 8.7% acetic acid, 0.001 M EDTA). Iodo-CMP can be seen as the spot with highest mobility that occurs in every lane labeled 4 (T2 RNase digests) and in many of the lanes labeled 3 (pancreatic RNase digests). This position was confirmed directly by running iodo-CMP obtained by RNase T2 digestion of iodo-poly(C) as in Figure 4A, lane 3.

Table II

Sequence assignment of RNase T1-resistant oligonucleotides from ^{125}I-labeled HeLa 5S RNA (see Figure 7D)

Oligonucleotide number	Sequence Assignment[b] (Forget and Weissman, 1967)	Secondary Analyses: Pancreatic RNase Digestion Products	Secondary Analyses: RNase U2 Digestion Products
1	CCAUACCACCCUG	AC,C	
2	AAUACCG	AC,C	CCG
3[a]		AC,C	CCG
4	AUCUCG	C	UCUCG
5	AACG, CUAAG mixture	AAC,C	CUAAG,AACG,ACG,CG
6	ACCG	AC,C	CCG
7	CCCG	C	CCCG
8	UACUUG	AC	CUUG
9	CCUG	C	CCUG
10	UCUG	C	UCUG
11	CUG, UCG mixture	C	CUG,UCG
12	CG	C	CG

[a]It was not possible to assign a sequence to oligonucleotide number 3. The intensity of this spot in the ^{125}I fingerprint (Figure 7C) was much reduced indicating that it may be from a minor species of 5S RNA which copurifies with the major species. This interpretation is supported by the presence of a minor spot in this same position in the ^{32}P fingerprint (Figure 7A). The sequence of this oligonucleotide appears to be related to that of spot number 2.

[b]Three C-containing oligonucleotides present in HeLa 5S RNA were not identified within the ^{125}I fingerprint shown in Figure 7C. These were spots e and g (CAG and CUU_{OH}) and one component of spot 4 (UCUACG), using the designations of Figure 7B and Table I. These anomalies are discussed in the text.

the ^{125}I fingerprint as did the analogous oligonucleotide in the ^{32}P fingerprint (compare Tables I and II and Figures 7B and 7D). Thus, we conclude that direct comparisons can be made between ^{32}P and ^{125}I fingerprints. In addition, the assignment of sequences to a number of oligonucleotides in Figure 8 permitted us to begin to compile a catalogue of mobilities for ^{125}I-labeled oligonucleotides at pH 1.9 (Figure 9).

There are two anomalies in the data obtained from secondary analyses of the ^{125}I-labeled oligonucleotides. Oligonucleotide number 3 does not contain the sequence CUAAG. Instead it contains a sequence which appears to be related to the sequence of oligonucleotide number 2: AAUACCG. It is not surprising that number 3 of the ^{125}I fingerprint is different from number 3 of the ^{32}P fingerprint since its mobility in the second dimension is slower than that of the ^{32}P-labeled oligonucleotide number 3. Furthermore, the intensity of this spot in the ^{125}I fingerprint was much reduced indicating that it may be from a minor species of 5S RNA with a slightly different sequence. This interpretation is supported by the presence of a minor spot in the analogous position of the ^{32}P fingerprint (Figure 7A). However, this still leaves the problem of locating the CUAAG which we have shown is present in HeLa 5S RNA and should be iodinated since it contains a cytidylate residue. Closer examination of the data for oligonucleotide number 5 from the ^{125}I fingerprint reveals that there is a contaminant of the AACG which has C as its only pancreatic RNase digestion product and has a mobility just a little slower than that of AACG when rerun with no prior digestion. These data suggest that CUAAG might run very close to AACG in the ^{125}I fingerprint.

The second problem is that neither CAG nor UCUACG were found in the ^{125}I fingerprint. Recent experiments by E. Pelle of The Rockefeller University have shown that CAG is sometimes present in the ^{125}I fingerprint in a position analogous to the one it occupies in

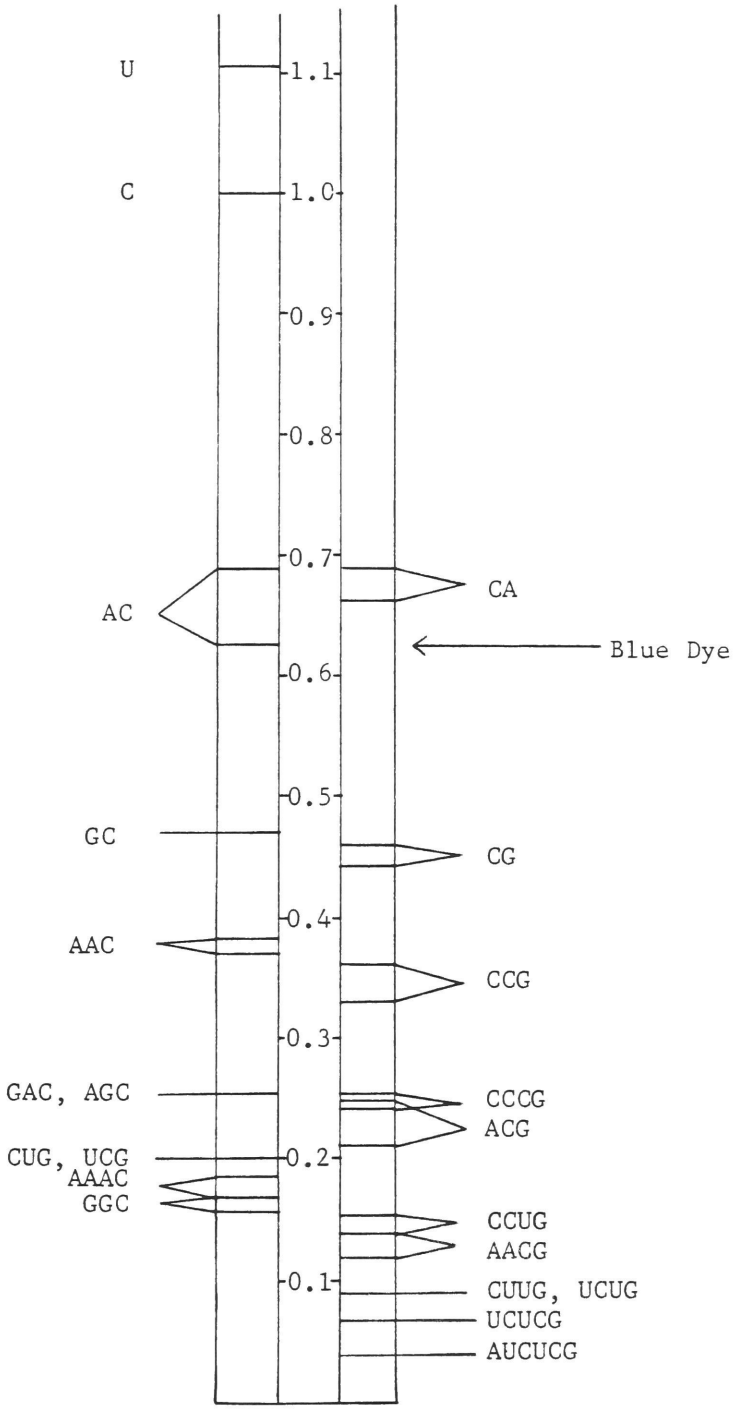


Figure 9

Mobilities of ^{125}I-labeled oligonucleotides at pH 1.9 (see Figures 7 and 8 and Tables II and III).

the ^{32}P fingerprint (personal communication, 1976; for an example of a ^{125}I-labeled 5S RNA in which CAG is clearly visible see Figure 3B). E. Pelle has also confirmed that CUAAG runs between the positions of oligonucleotides number 4 and 5 and that it is frequently a contaminant of spot number 5 (AACG). However, he finds no evidence of the presence of UCUACG except possibly as a trace contaminant of spot number 4. In the ^{32}P fingerprint, number 4 corresponds to a mixture of two moles of AUCUCG and one mole of UCUACG. Sequence isomers such as these almost always run as a single spot in such fingerprints where the second dimension separation is achieved by ascending homochromatography.

It is most probable that if UCUACG were present, as a ^{125}I-labeled oligonucleotide, it would be detected as a mixture with AUCUCG in spot number 4 by the presence of CG among the RNase U2 digestion products. There are two possible explanations for the absence of this oligonucleotide from the ^{125}I fingerprint. It is possible that the major species of 5S RNA in the HeLa cell cultures from which the unlabeled RNA was isolated was a different species from the major one present in both the HeLa cells from which the ^{32}P-labeled 5S RNA was isolated and the KB cells studied by Forget and Weissman (1967). A second possibility is that the two cytidylate residues in UCUACG are somehow protected from iodination. Forget and Weissman (1967) have proposed that 5S RNA could contain secondary structure including base pairs between residues 1 through 9 and 118 through 110, thus joining the 5'-end to the 3'-end of the molecule as in the cloverleaf model for tRNA. UCUACG represents residue 2 through 7. Several studies have suggested that C's within helical regions have a much reduced rate of iodination in comparison to those in single-stranded regions (Commerford, 1971; Varricchio and Prensky, in prep.). In further support of this idea, Forget and Weissman propose that there is a second region of hydrogen bonding within KB cell 5S. The only distinctive T1-resistant oligonucleotide in this second region is CCCG (spot number 7 in the ^{125}I fingerprint, Figure 7B). If all cytidylate residues in

the 5S molecule have an equal chance of being iodinated, we would expect CCCG to have three times more ^{125}I than the amount found in AACG (spot number 5). A visual inspection of the ^{125}I fingerprint shown in Figure 3B indicates that this is not the case. In fact, AACG appears to be more heavily labeled than CCCG. This observation, coupled with the fact that the oligonucleotide UCUACG can not be detected at all, suggests that these regions have reduced accessibility to the iodination reaction.

Because of these indications that RNA regions with double helical structure might be less readily iodinated under our reaction conditions, it was necessary to define more carefully both the capabilities and the drawbacks of the iodination reaction. In order to do this we asked two questions: (i) Does the temperature of iodination radically affect the fingerprint) (ii) Is there any intrinsic bias involved in iodination (for example, nearest neighbour cytidylate residues can never both be iodinated) which would prevent a random distribution of iodo-C residues? Once again, HeLa 5S RNA was used as the model RNA species for these studies.

Effect of Temperature on Iodination

As documented by Commerford (1971) single-stranded DNA is more susceptible to iodination than double-stranded, and the rate of iodination increases as temperature increases. When taken together with the tendency of nucleic acid secondary structure to become less stable with increased temperature, these properties could cause different patterns of iodination under different conditions. Because of technical difficulties involved in handling both the tiny volumes and the relatively high amounts of radioactivity used in these reactions, it is extremely difficult to achieve precise control of the temperature and duration of the reaction. Thus, it would be most convenient if we could find a reasonably broad range of temperatures of iodination which would yield essentially the same fingerprints. HeLa 5S RNA

iodinated at either 40° C or 70° C was digested with RNase T1 and fingerprinted (Figures 10A and 10B). These fingerprints were compared with that obtained after iodination under conventional conditions (60° C; Prensky, 1975) (Figure 3B). No significant changes in the distribution of radioactivity to the various oligonucleotides can be detected. More extensive studies on the effect of temperature during iodination upon the distribution of ^{125}I within the cytidylate residues of HeLa 5S RNA have been carried out by Jacobson, Housman, and Prensky (1973) with essentially the same results. However, in some cases (one is shown in Figure 11) the T1 fingerprint of HeLa 5S demonstrates a much more representative amount of label in both the CAG spot and the CCCG spot. The reason for this still remains a mystery.

A clue to such variation may be found in some recent work by Varrichio and Prensky (manuscript in preparation). In the case of E. coli formylmethionine transfer RNA (an RNA species with substantially more secondary structure than the ribosomal 5S RNA), these authors observed a profound effect of the presence of low concentrations of magnesium ions on the accessibility of various regions of the tRNA to the iodination reaction. It could be that RNA samples differing from each other by the presence or absence of low amounts of cations could lead to the differences observed between fingerprints such as Figure 3B, 7C, 10B and 11A.

The main point to make for our purposes, however, is that the fingerprints obtained after iodination at these three widely spaced temperatures are remarkably similar. Each one can be recognized readily as a HeLa 5S T1 fingerprint. Thus, for comparative fingerprinting studies of single-stranded RNA's, the choice of temperature of iodination should not be crucial within a fairly broad range. When previously uncharacterized RNA species are studied, however, they should be iodinated under conventional conditions and also at elevated temperatures. Examination of subsequent T1 fingerprints would indi-

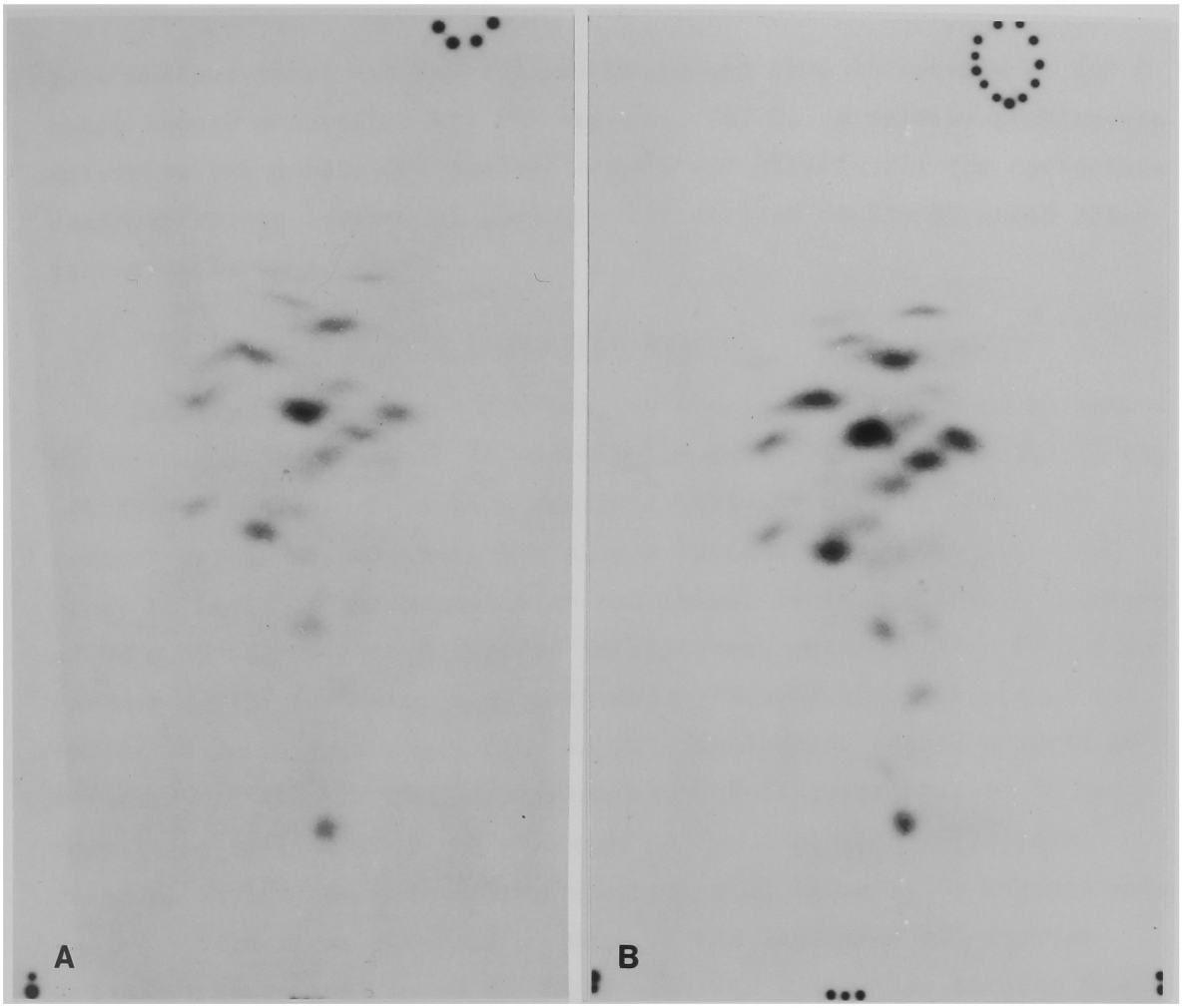


Figure 10

The effect of temperature upon iodination of HeLa 5S RNA. Iodination was carried out at 40° C and 70° C and the resulting ^{125}I-labeled RNA yielded T1 fingerprints shown in (A) and (B) respectively. These fingerprints can be compared with the T1 fingerprint of ^{125}I-labeled HeLa 5S RNA in which iodination was carried out in the conventional manner at 60° C in Figure 3B.

cate whether there are any regions protected from iodination at 60° C which become accessible at, for example, 75° C. A related problem--to determine the conditions needed to iodinate efficiently the cytidylate residues within regions of biologically derived double-stranded RNA--is currently under study.

Is Iodination Random?

Ideally, for sequence studies, we should have access to methods of iodination that result in essentially random labeling of all of the cytidylate residues to a high specific activity without bias with respect to the nearest neighbor of the residue to be modified. In order to test for the presence of such bias, we have prepared a series of HeLa 5S RNA samples iodinated to different extents (1 to 28%) (see Prensky (1975) for details of methods). Because the presence of one iodinated cytidylate residue in an oligonucleotide caused a shift to a faster mobility in the first dimension of fingerprints, it is not surprising to find that the presence of two iodinated cytidylate residues within the same oligonucleotide will cause it to migrate even faster in the first dimension. Most of the iodinated RNA species discussed up to this point have had specific activities ranging from 10^6 to 10^8 cpm per microgram. This corresponds to iodination of from .048 to 4.8% of the cytidylate residues (see Appendix I for the method of calculation of percentage iodination from the specific activity of labeling). Even at a specific activity of 10^8 cpm per microgram, the probability of finding two iodinated cytidylate residues within one oligonucleotide is very small. However, if the percentage iodination is increased so that about 25% of the C's become iodinated, it is clear that oligonucleotides such as the largest T1-resistant oligonucleotide of HeLa 5S RNA (CCAUACCACCCUG) should contain anywhere from one to seven iodo-C residues. Furthermore, if iodination is random, the distribution of oligonucleotides containing various extents of iodination should follow a Binomial distribution. Figures 11A and 11B show two fingerprints of HeLa 5S RNA iodinated to 1% and 28% of the

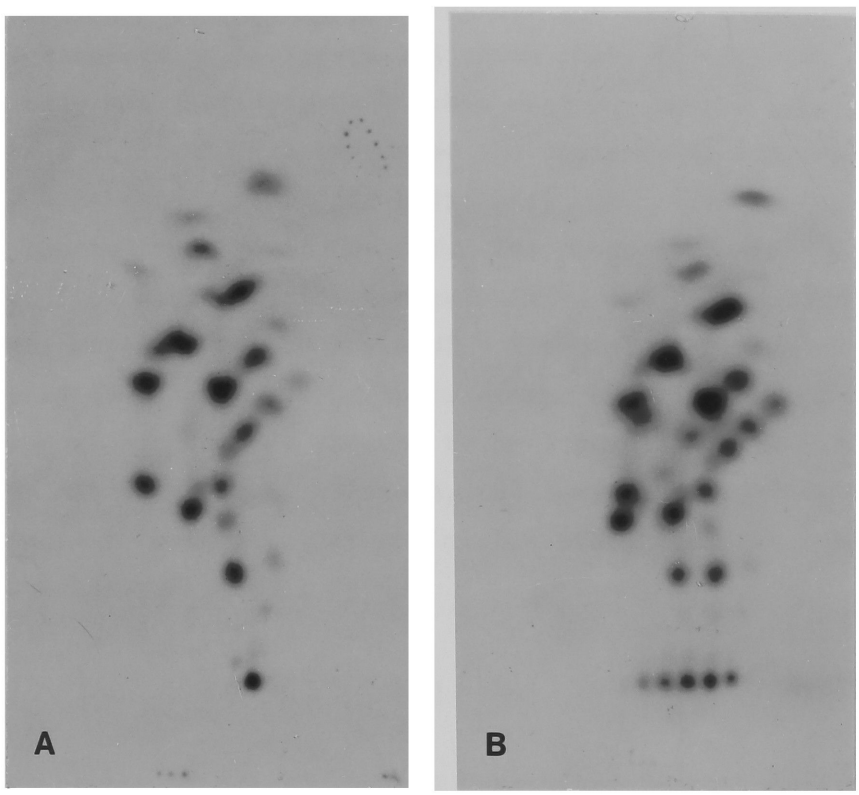


Figure 11

Increased levels of iodination of HeLa 5S RNA. Two aliquots of HeLa 5S RNA were iodinated to specific activities corresponding to 1% and 28% of the cytidylate residues. (A) T1 fingerprint of 10^6 cpm of HeLa 5S RNA with 1% of the C's iodinated. (B) T1 fingerprint of 10^7 cpm of HeLa 5S RNA with 28% of the C's iodinated. (See the text and Appendix II for an analysis showing that iodination is random.)

C's respectively. In Figure 11B the five evenly spaced spots arranged in a horizontal row near the bottom of the fingerprint represent the oligonucleotide CCAUACCACCCUG with 1, 2, 3, 4, or 5 of its C's iodinated (reading from right to left within the fingerprint). Comparison with the fingerprint in Figure 11A shows that, in the case of 1% iodination, only one spot appears in this region. Longer exposure of the fingerprint in Figure 11B indicates the presence of a sixth and even possibly a seventh spot in the horizontal row. These spots were cut out and counted in a gamma counter. The proportion of ^{125}I gamma cpm recovered from each position was calculated and the results compared with those expected on the basis of a Binomial distribution of iodination events (see Table III). Fortunately, this 13-nucleotide fragment is located within what is thought to be a purely single-stranded region of the 5S RNA (Forget and Weissman, 1967) so that its iodination should not be subject to temperature and cation effects of the type discussed in the preceding section.)

Table III

Comparison of Experimentally Determined Distribution of Iodination and a Random Distribution Calculated from the Binomial Distribution

"n" Number of iodo-C's in oligonucleotide	Gamma counts per minute	Percentage of ^{125}I recovered from the spot with "n"C's labeled	Percentage expected based on Binomial Distribution (see Appendix II)
1	17,660	13.7	13.92
2	34,406	27.0	32.49
3	33,798	26.5	31.65
4	23,627	18.5	16.38
5	13,430	10.5	4.77
6	3,501	2.7	0.74
7	781	0.6	0.05

Because the distribution of ^{125}I does not diverge markedly from the Binomial Distribution, we conclude that there is no bias in iodination with respect to preferred nearest neighbor and further, that as the extent of iodination is increased, there is no hindrance to the iodination of all the C's present. Thus, it should not be difficult to obtain RNA labeled to very high specific activity with nevertheless a random distribution of the iodo-C's throughout the molecule for the purposes of RNA sequence determinations.

Approaches to Sequence Analysis of Iodinated RNA

It will be intrinsically more difficult to carry out sequence determination studies with ^{125}I-labeled RNA than with ^{32}P-labeled RNA since, in the former case, only one of the four bases can be labeled while in the case of ^{32}P all four can be labeled if so desired. However, one rationale that could be applied to iodinated RNA's has been investigated. Spleen phosphodiesterase is a commercially available exonuclease capable of digesting RNA in a stepwise fashion in a 5' to 3' direction. If we obtain iodinated oligonucleotides that have iodo-C at their 3'-termini, digest them with spleen phosphodiesterase, and analyze the products by two-dimensional fingerprint analysis, it should be possible to determine, from the change in mobility with each subsequent loss of a base, what the sequence of the oligonucleotide is. This approach was originated by Ling (1972) who studied ^{32}P-labeled polypyrimidine tracts from bacteriophage DNA. Of course any such approach would have to be supplemented by other analyses in order to obtain conclusive evidence for the presence of a particular sequence. Figure 12 illustrates a preliminary attempt to apply this rationale. The oligonucleotide AAUACCG has been eluted from a T1 fingerprint of ^{125}I-labeled HeLa 5S RNA, digested partially with spleen phosphodiesterase, and fingerprinted (Figure 12A). A schematic drawing of the proposed partial products of digestion are indicated in Figure 12B. Thus, this exonuclease appears to work just as well

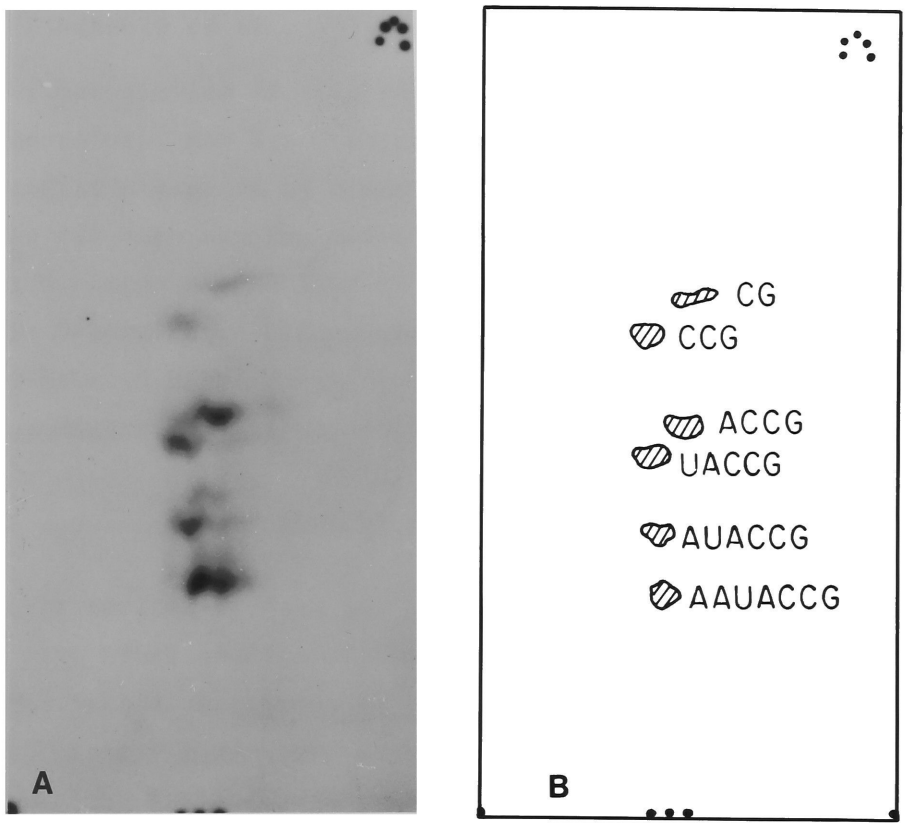


Figure 12

Fingerprint of a spleen phosphodiesterase digest of the iodinated oligonucleotide AAUACCG from HeLa 5S RNA. The ^{125}I-labeled oligonucleotide was incubated in 2 microliters of 1 mg/ml enzyme in buffer containing 0.1 M ammonium acetate (pH 5.5) - 0.002 M EDTA - 0.05% (v/v) Tween 80 (Atlas Chemical Industries) for 0-100 minutes and fingerprinted using the conventional techniques: (A) Fingerprint of the partially digested oligonucleotide. (B) Schematic drawing indicating the proposed partial cleavage events (Barrell, 1971; Ling, 1972).

on iodinated oligonucleotides as it does on ^{32}P-labeled oligonucleotides (Pieczenik et al., 1974).

Further studies of this kind have been conducted by S. Legon of this laboratory, who has combined partial spleen exonuclease digestion with partial digestion by other nucleases in order to obtain sequences from the ribosome binding sites of rabbit globin mRNA. During these studies S. Legon showed that it is possible to obtain the expected specific C-terminated oligonucleotides from a pancreatic RNase digest of ^{125}I-labeled RNA containing U residues chemically blocked from nuclease digestion (Gilham, 1962; Barrell, 1971).

SUMMARY AND CONCLUSIONS

While earlier studies had shown the usefulness of iodinated RNA for *in situ* hybridization (Prensky et al., 1973), it remained to be shown that the distribution of the iodinated residues within the RNA was sufficiently random and reproducible for it to be employed in fingerprinting and sequencing analyses. The studies presented in this chapter have established the following points.

1. Fingerprints of iodinated bacteriophage f2 RNA and human 5S RNA resemble those of their non-iodinated counterparts both in complexity and in specific pattern.

2. Base composition analyses using RNase T2 on both poly(C) and biological RNA's has shown that over 95% of the ^{125}I label is present as iodo-C.

3. Alkaline hydrolysis of iodinated RNA under conditions of conventional base composition analysis leads either to deamination or de-iodination of most iodo-C residues and production of either iodo-U or non-iodinated C.

4. The mobility of iodinated oligonucleotides differs from the non-iodinated counterparts in all electrophoretic systems tested, whereas they are not significantly different when subjected to

ascending homochromatography. RNA doubly labeled in its cytidylate residues with ^{14}C and ^{125}I was used to show that a two-dimensional fractionation using a homochromatographic second dimension gives the best comparison between iodinated and non-iodinated RNA. Such ^{14}C and ^{125}I fingerprints were found to differ by a systematic shift to a higher first dimension mobility on the part of the iodinated oligonucleotides.

5. Comparative enzymatic analysis of oligonucleotides from ^{32}P- and ^{125}I-labeled HeLa 5S RNA demonstrated that the ribonucleases normally used for RNA sequence analysis (RNases T1, T2, U2, pancreatic RNase and spleen phosphodiesterase) cleave iodinated RNA with unaltered specificity.

6. A catalogue of electrophoretic mobilities at pH 1.9 of iodinated oligonucleotides has been compiled.

7. Analysis of the distribution of ^{125}I in HeLa 5S RNA containing 1% or 28% iodo-C residues has established the essentially random nature of the iodination reactions used here when applied to single-stranded RNA species.

These studies have established a new approach for RNA sequence analysis, which has already been used to investigate properties of eukaryotic RNA species previously inaccessible to such techniques.

CHAPTER II

STUDIES ON VIROID RNA

The spindle tuber disease of potato has been recognized as an economically important disease for many years (Schultz and Folsom, 1923). Symptoms including stunted foliage (Figure 13A and 13B) and disfigured tubers can cause crop reductions of 64% in affected plants (Singh et al., 1970; Singh et al., 1971). The disease is thought to be transmitted horizontally by means of harvesting equipment (Goss, 1926) while vertical transmission occurs with high frequency through the true seed (Singh, 1970; Singh and Finnie, 1973).

Thus, as characterized by both its symptoms and natural means of transmission, spindle tuber disease of potato is similar to many plant virus diseases (Diener and Raymer, 1971). On the other hand, no virus-like particles could be found in extracts of diseased tissue (Diener and Raymer, 1969; Diener, 1971a). In fact, the infectious agent seemed to possess the properties of a free nucleic acid (Raymer and Diener, 1969). This surprising discovery was followed by a series of studies by Diener and colleagues which led, three years later, to the isolation of a low molecular weight species of RNA, detectable only in diseased plants, whose electrophoretic mobility in 20% polyacrylamide gels coincided exactly with that of the mysterious infectious agent (Diener, 1972).

Because the pathogen was infectious as free RNA and did not seem to contain sufficient genetic information for self-replication ("genome" size was estimated at $2.5 \times 10^4 - 1.1 \times 10^5$ daltons), it seemed likely that the infectious entity depended for its replication upon biosynthetic systems present in the uninoculated plant (Diener, 1971b).

The dramatic differences observed between this infectious free RNA molecule and the conventional virus systems led Diener to propose that the term "viroid" be adopted to designate potato spindle tuber "virus"

Figure 13A

Comparison of healthy and PSTV-infected tomato plants. Seedlings of Lycopersicon esculentum cv. Rutgers were inoculated with PSTV at the four-leaf stage and grown in the greenhouse for three weeks. The plant on the left hand side is the PSTV-infected specimen. The healthy plant on the right is the same age.

Figure 13B

Comparison of leaves removed from the PSTV-infected plant shown in Figure 13A with those removed from an uninfected plant of the same age.

RNA and all other RNA's with similar properties (Diener, 1971b). Previously the term "viroid" had been defined as "any prophylactic vaccine" (Stedman's Medical Dictionary, 1961) or "any biological specific used in immunization" (Dorland's Illustrated Medical Dictionary, 1965). In 1946 the term was redefined by Altenburg as hypothetical, ultramicroscopic organisms that are useful symbionts, occur universally within cells of larger organisms, and are capable, by mutation, of giving rise to viruses. These two earlier meanings of the word "viroid" were judged to be obsolete (Diener, 1973) at the time it was defined for a third time (Diener, 1971b). With the exception of Singh and colleagues who in 1973 coined the term "metavirus" (meaning "after" or "beyond" viruses) to refer to this same class of plant pathogens (Bagnall et al., 1973), other workers in the area have accepted Diener's terminology.

Other Diseases Caused by Viroids

Since Diener demonstrated the existence of the potato spindle tuber viroid (PSTV) three other plant diseases have been linked directly with the presence of a viroid: (i) citrus exocortis viroid (CEV) (Semancik and Weathers, 1972a; Sänger, 1972), (ii) chrysanthemum stunt viroid (CSV) (Diener and Lawson, 1973; Hollings and Stone, 1973) and (iii) chrysanthemum chlorotic mottle viroid (ChCMV) (Romaine and Hörst, 1975). In all of these cases, low molecular weight RNA has been isolated from diseased plants, purified by polyacrylamide gel electrophoresis, and shown to be infectious (these represent the minimum criteria necessary to establish the presence of a viroid).

Two additional examples of potential viroid diseases have been described. Van Dorst and Peters (1974) have reported the existence of unpublished experiments proving that the pale fruit disease of cucumber is caused by a viroid (CPFV). Finally, Randles (1975) suspects that cadang-cadang (a disease of coconut palms of enormous economic importance in the Philippines) is caused by a viroid. In this case, a low molecular weight RNA of the appropriate size present only in diseased

tissue has been isolated. Unfortunately, it has not yet been possible to demonstrate infectivity of this new RNA species.

Evidence for the Existence of Viroids

There is no question as to the existence of the plant diseases mentioned above, nor would anyone question the fact that infectious entities capable of transmitting these diseases also exist. However, the prospect that such pathogens might be composed entirely of RNA of low molecular weight (the estimates range from about 250 to 350 nucleotides), has been seriously questioned by several respected colleagues.

The most powerful evidence available for the existence of a pathogen with viroid properties comes from the experiments carried out by Diener (1972) to correlate infectivity with a band of UV-absorbing material present only in RNA preparations from diseased tissue. Because these results have been of central importance in moving the study of viroids from a study of the phenomenology of plant disease into the arena of the molecular biologist they will be presented in some detail here. RNA was purified from both healthy and PSTV-infected tomato plants by a procedure including, among other steps, organic extraction, DNase treatment, removal of material which precipitates with 2M LiCl treatment, and fractionation of the resulting RNA on 20% cylindrical polyacrylamide gels. After electrophoresis sufficient to move 5S RNA to the bottom of the gel, the optical density profile at 260 nm was recorded. Low molecular weight RNA from healthy plants was found to contain three well-resolved unidentified RNA species with electrophoretic mobilities slower than that of 5S. PSTV-infected plants contained an additional UV-absorbing species situated between two of these host-specific RNA species. When the gel was sliced and assayed for the presence of PSTV, a peak of infectivity was observed whose position corresponded exactly with that of the UV-absorbing component present only in PSTV-infected plant RNA.

Two formally acceptable explanations for these results need to be considered. Either the major component of the new gel band is the

pathogenic agent itself, or else it is a host-specified product that is over-produced in the presence of a viroid and, coincidentally, has precisely the same electrophoretic mobility as the viroid. The probability that this latter suggestion is true is obviously extremely low. Nevertheless, direct evidence will be required to eliminate it. Meanwhile, the assumption will be made that the former, more likely explanation is the correct one.

Properties of Viroids

PSTV and CEV are the most well-characterized of the viroids. In each case, the infectious agent has been identified as a free RNA species of low molecular weight, with size estimates ranging from 250 to 350 nucleotides (Diener, 1971b; Diener and Smith, 1973; Semancik and Weathers, 1972a; Semancik et al., 1973b; Sogo et al., 1973). Recent evidence indicates that infectious RNA 80 to 100 nucleotides long is also recovered in the case of infection of Scopolia sinensis by PSTV (Singh et al., 1974).

Detailed studies by Diener and co-workers have shown that PSTV replicates in susceptible host species without the assistance of a helper virus (Diener, 1971b; Diener et al., 1972). Because of their small size and inability to act as messenger RNA in several cell-free protein synthesizing systems (Davies et al., 1974; Hall et al., 1974) it seems unlikely that viroids carry information for the specification of a complete replicase in their genome. On the other hand, viroid RNA could, formally speaking, represent the negative strand of a messenger RNA which, when transcribed, would synthesize, for example, a replicase subunit. The possibility that the viroid genomes have higher complexity and consist of a number of different RNA fragments of similar size was investigated by Diener et al. (1974) who measured the ultraviolet light sensitivity of PSTV and by Semancik et al. (1973b) who studied the inactivation of CEV by ionizing radiation. Both groups concluded that the size of the biologically active unit was approximately equal to that determined by polyacrylamide gel electrophoretic mobility.

PSTV and CEV replicate in a wide variety of plants but cause disease symptoms in only a few; no host has yet been found which responds differently to the two agents (O'Brien and Raymer, 1964; Singh and O'Brien, 1970; Singh, 1971; Diener et al., 1972; Semancik and Weathers, 1972b, c; Singh, 1973). Bioassay has shown that these RNA's produce identical disease symptoms across a common host range leading to the suggestion that they may be independent isolates of the same pathogen (Semancik and Weathers, 1972c; Semancik et al., 1973a; Singh and Clark, 1973). Others have cited a number of other features which are held in common in order to lend additional credence to this conjecture: (i) the thermal stability of both RNA's resembles that of tRNA more closely than that of double-stranded RNA (Diener, 1972; Semancik et al., 1975); (ii) the infectivity of both agents is resistant to heat treatment (Singh and Bagnall, 1968; Semancik et al., 1973b); and (iii) the infectious RNA of partially purified PSTV has been observed to migrate approximately like that of CEV upon electrophoresis in 5% polyacrylamide gels (Semancik et al., 1973a).

The work presented in Chapter I demonstrates that RNA fingerprinting and sequencing techniques originally developed using ^{32}P-labeled RNA species (Brownlee and Sanger, 1969; Barrell, 1971) can be extended to include applications to RNA labeled in vitro with ^{125}I. We have therefore subjected ^{125}I-labeled RNA of PSTV and CEV to two-dimensional fingerprinting analysis following specific nuclease digestion to determine (i) whether these RNA's have complexities consistent with their reported sizes and (ii) whether PSTV and CEV RNA are identical in primary structure. A description of these studies, based partially upon a paper by E. Dickson, W. Prensky, and H. D. Robertson (1975) is presented in this chapter.

A number of profound biological effects have been linked with the presence of double-stranded RNA (Robertson and Hunter, 1975). The possibility that PSTV and CEV might contain extended regions of perfect double-stranded RNA was investigated by studying the effect of

digestion by E. coli Ribonuclease III, an enzyme that digests double-stranded RNA (Robertson, Webster and Zinder, 1968).

The Viroid Genotype is the Viroid Phenotype

As will be shown in this chapter, the only unique feature a particular viroid has that allows it to be distinguished chemically from any other viroid is its RNA sequence. Until a direct assay, such as fingerprinting of iodinated RNA, is applied to viroids isolated from various plants, there will be no way of knowing whether the viroid isolated from a group of diseased plants is the same as the viroid used to inoculate the plants. The requirement for using such an assay becomes more obvious in light of observations by D. Peters (personal communication, 1975), that, in at least some cases, the viroid appears to change its potential for pathogenicity upon passage through a particular host. Thus, the possibility that a viroid can be modified by its host arises. Furthermore, replication of a viroid in its host plant does not always cause disease symptoms (e.g. PSTV in tobacco; Diener et al., 1972) and certain mild (Singh et al., 1970) or symptomless (Hörst, 1975) viroid strains can protect the plant from subsequent disease caused by pathogenic strains. In order to sort out the relationships among these RNA species and the relationship, if any, to the host genomes, it will be necessary to carry out sequence analysis of a large number of these RNA molecules. One way to embark upon this long-term project is to begin compiling a file of fingerprints of the iodinated RNA of any available viroids so that they may be used for future identification. Then, if host modification does occur, it will be detectable by a change in the fingerprint. The results of two-dimensional fingerprint analysis of iodinated RNA from CSV and viroid-sized RNA from coconut palms with the cadang-cadang disease are presented in this chapter.

Since the long-term goals of understanding relationships between viroids and their hosts can only be met if it is possible to characterize a reasonably large sample of viroid RNA's after passage through

a variety of their hosts, and since, by current technology, substantial amounts of plant material are required to yield microgram quantities of viroid RNA, we have been developing techniques for isolation of highly purified viroid RNA on a micro scale in a form that can be iodinated. Preliminary results of these studies will be presented.

MATERIALS AND METHODS

Viroid RNA: PSTV RNA isolated from the tomato plant, Lycopersicon esculentum (cv. Rutgers), was kindly provided by Dr. T. O. Diener (U.S.D.A., Beltsville, Md.). CEV RNA isolated from Gynura aurantiaca was the generous gift of Dr. J. S. Semancik (Univ. of California, Riverside, Cal.). CSV and the coconut palm RNA's were isolated by Dr. T. O. Diener (in collaboration with Dr. Randles in the latter case), and were sent to us for analysis by these investigators.

Iodination, Fingerprinting, and Secondary Enzymatic Analysis of RNA: see Chapter I.

Polyacrylamide Gel Electrophoresis: All slab gels (20 x 40 x 0.3 cm) had a ratio of acrylamide (Canalco) to bisacrylamide (Canalco) of 20 to 1. The non-denaturing gels were run at 4° C in buffer containing 10.8 grams of Tris base, 0.93 grams of disodium EDTA (ethylenediaminetetraacetic acid) and 5.5 grams boric acid per liter (pH 8.3) (Peacock and Dingman, 1968). For denaturing gels, the appropriate amounts of acrylamide and bisacrylamide were dissolved directly in formamide (99%, MC/B) that had been previously deionized by treatment with mixed bed resin (Bio-Rad) and buffered by addition of mono- and di-basic sodium phosphate to obtain pH 7.0 as described by Duesberg and Vogt (1973). Formamide containing gels were run at room temperature in 0.04 M sodium phosphate buffer, pH 7.0. Polymerization was brought about by the addition of ammonium persulphate and TEMED (N,N,N',N'-tetramethylethylenediamine; Eastman Kodak Co.) each to a final concentration of 0.075%. Details of individual runs are included in the figure legends. The dye marker, bromphenol blue, was

included in each sample. RNA samples (one to two micrograms each) were dissolved either in water with 10% sucrose or in buffered deionized formamide with 10% glycerol. At the end of each run, the gel was stained for 15 minutes in a solution of methylene blue (0.2%) containing 54 grams sodium acetate and 9.2 milliliters glacial acetic acid per liter, and then destained in water.

Purification of RNA by Chromatography on Whatman CF11 Cellulose: see Chapter I Methods and Materials. This procedure is used here mainly to separate the RNA from contaminating materials such as Tris and acrylamide prior to iodination, or to separate ^{125}I-labeled RNA from free ^{125}I anions following iodination.

HCl Precipitation of RNA: This procedure was carried out on a variety of RNA samples, usually to remove impurities that would interfere either with the iodination reaction or with the fingerprint analysis. RNA samples were dissolved in 0.3 ml of water (containing 10 micrograms of carrier RNA if an iodinated sample is to be precipitated) and placed in 12 ml Pyrex siliconized Sorvall centrifuge tubes on ice. Six microliters of concentrated hydrochloric acid were added to each tube and the tubes swirled on ice to mix immediately and thoroughly. Two minutes later, the tubes were transferred to an SS34 rotor of a Sorvall centrifuge at 4° C and centrifuged at 15,000 rpm for 40 minutes. The supernatants were removed carefully with Pasteur pipets and the pellets (not visible) dried down for 3 to 5 minutes in a vacuum dessicator. The pellets were then resuspended in two 15 microliter aliquots of 30% TEC. These washes were frozen and lyophilized in a vacuum dessicator. After two cycles of resuspending in water and drying down, the RNA was considered ready for analysis.

RESULTS

Complexity of Viroid RNA

Before the studies reported here were carried out, there was a formal possibility that viroid RNA preparations were complex mixtures

of many different RNA species. Comparison of the two-dimensional fingerprints of RNase T1 digests of ^{125}I-labeled PSTV, HeLa cell 5S rRNA, and duck globin mRNA (Figure 14) shows that, while PSTV RNA is more complex than the 5S RNA, it is substantially less complex than the globin mRNA. Thus, the sequence complexity of the major component of this viroid RNA preparation is consistent with the size estimate of 250 to 350 nucleotides.

Comparative Studies of PSTV and CEV

Polyacrylamide Gel Electrophoresis: To compare the mobilities of PSTV and CEV under highly resolved conditions we have used slab gels (20 x 40 x 0.3 cm) with adjacent slots containing PSTV, PSTV and CEV mixed, and CEV alone (shown in Figure 15 from left to right respectively). Compositions of the five different gel systems employed for these studies are described in the legend to Figure 15. In 5% polyacrylamide gels in borate buffer, pH 8.3 (Figure 15A), each viroid migrates as a single band with PSTV moving about 3% further than CEV. Previously, Semancik et al. (1973a) used 5% polyacrylamide gels of lower resolving ability (8 cm cylindrical gels; Tris-acetate-EDTA, pH 7.2, with 0.2% sodium lauryl sulphate) and found that infectious RNA's of PSTV and CEV had approximately the same mobilities, although in these studies a 3% mobility difference such as that shown in Figure 15A would not have been detected.

When the percentage of the gel is increased to 10% (Figure 15B), the PSTV again runs as a single band slightly faster than CEV, while the CEV migrates as two components--the minor one running just barely faster than the major component, but still slower than PSTV. (This latter component is very difficult to see in the photographic representation of the stained gel.)

When the percent acrylamide is further increased to 15% (Figure 15C) both PSTV and CEV demonstrate two well separated components. Furthermore, the proportion of each viroid RNA present in each component is different in the mixture of the viroids (Figure 15C, middle lane)

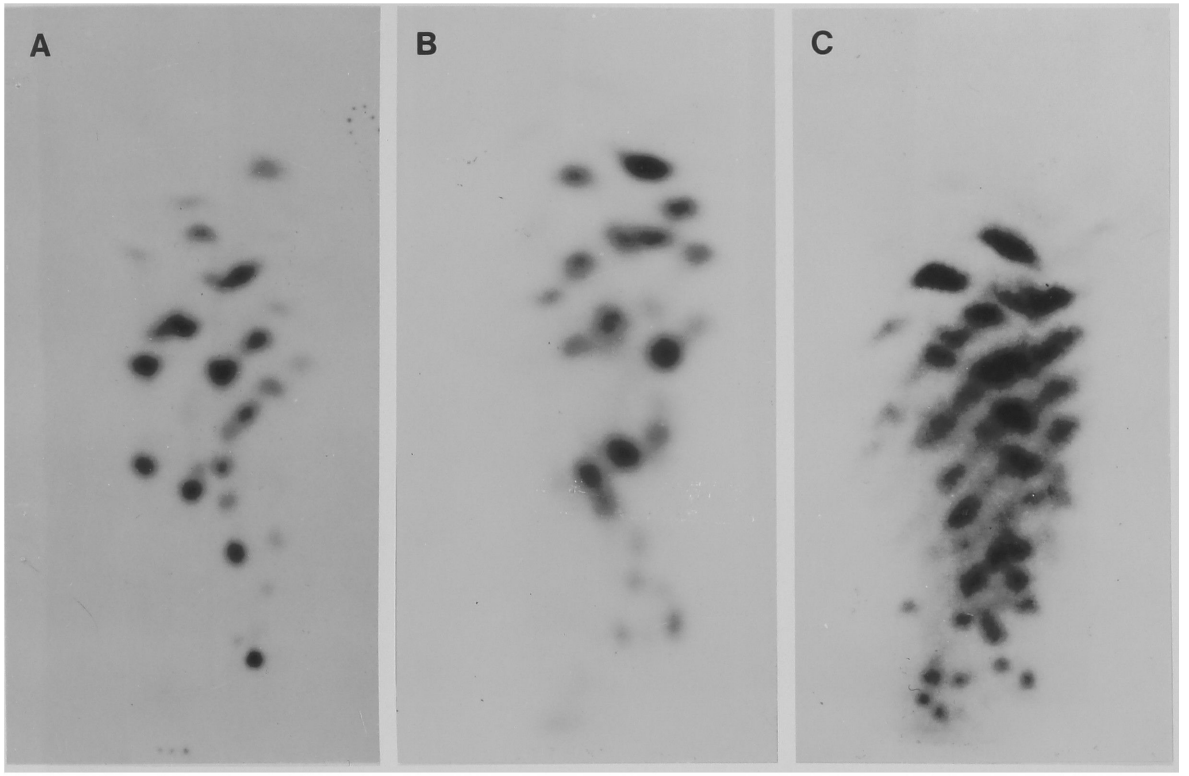


Figure 14

Comparison of the complexity of PSTV with that of HeLa 5S RNA and duck globin mRNA. Samples of HeLa 5S RNA, PSTV and globin mRNA were iodinated and their T1 digests fingerprinted according to the methods outlined in Chapter I. (A) HeLa 5S T1 fingerprint. (B) PSTV T1 fingerprint. (C) Duck globin mRNA T1 fingerprint. The upper regions of all T1 fingerprints contain the same oligonucleotides (e.g. CG, CUG, ACG, etc.). The lower portions contain the longer oligonucleotides which uniquely characterize each RNA species. The size of the longest RNase T1 product increases with an increase in the complexity of the RNA in question (Jeppesen, 1971).

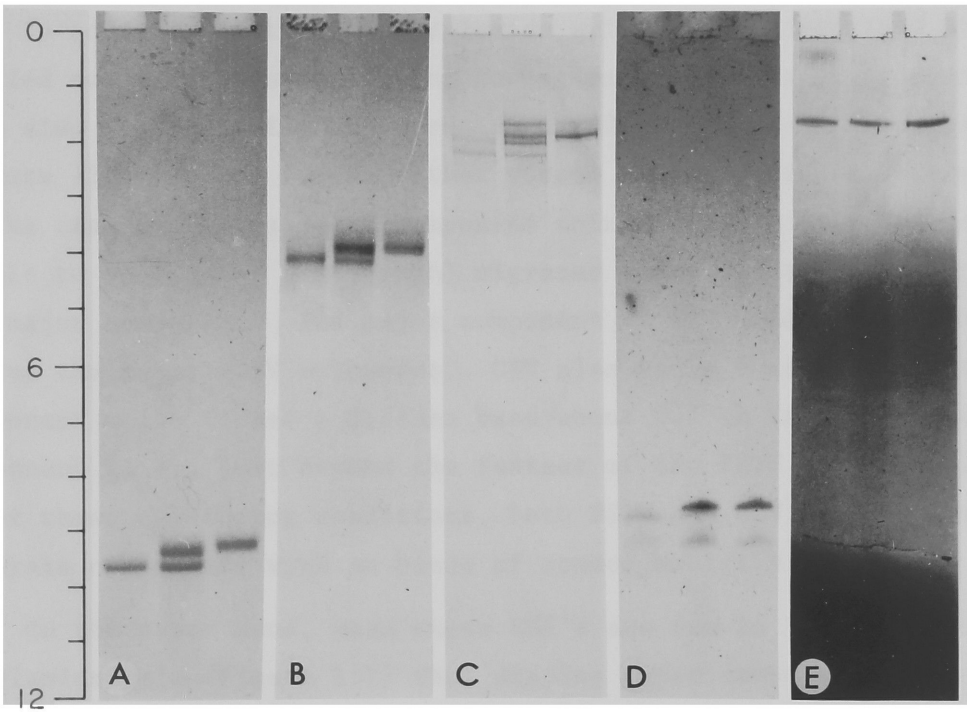


Figure 15

Polyacrylamide gel electrophoretic analyses of viroid RNA. PSTV and CEV RNA's have been run in five different polyacrylamide gel systems and are shown here side by side for direct comparison. The distance from the origin (cm) is indicated by the scale on the left side of the figure. In each system three samples were run, always in the same order from left to right: PSTV (1 microgram); PSTV (1 microgram) mixed with CEV (1 microgram); CEV (1 microgram). After running, the gels were stained with methylene blue, destained in water, and photographed. (A) 5% acrylamide gel in borate buffer, pH 8.3 (4° C, 10 volts per cm, 12 hours). (B) 10% acrylamide gel in borate buffer, pH 8.3 (4° C, 10 volts per cm, 24 hours). (C) 15% acrylamide gel in borate buffer, pH 8.3 (4° C, 10 volts per cm, 56 hours). (D) 5% acrylamide gel in formamide buffered with sodium phosphate pH 7.0 (5 volts per cm, room temperature, 12 hours). (E) 10% acrylamide gel in formamide buffered with sodium phosphate, pH 7.0 (5 volts per cm, room temperature, 24 hours). See Materials and Methods for additional details of gel preparation.

than when these RNA's are run separately (Figure 15C, outside lanes). The quantitative differences seen here remain to be explained.

In addition to the polyacrylamide gel analyses described above and carried out under non-denaturing conditions, two denaturing systems were also studied. Electrophoresis in 98% formamide-5% acrylamide gels (Figure 15D) did not reduce either viroid RNA to a single component. In the case of PSTV a minor component which formed a diffuse band (difficult to see in the photograph) migrated about 0.4 cm farther than the major component. The major component of CEV banded just 0.1 cm behind the major PSTV component. CEV also had a faster moving minor component which formed a diffuse band about 0.7 cm from the major CEV component (i.e., just beyond the fastest of the PSTV components). Thus, under these denaturing conditions, both PSTV and CEV behave as two separate components with no bands of common mobility.

On the other hand, when these RNA's are run in 98% formamide-10% acrylamide gels (Figure 15E) they display major components of identical mobility. In addition, a minor diffuse band can be seen about one quarter of the way between the origin and the major band in the case of PSTV alone (left hand lane). No such component can be detected in either the middle lane where the two viroids are run as a mixture, or in the right hand lane where CEV is run alone.

Comparisons Based on Electrophoretic Mobility: RNA species of identical primary structure which are isolated by identical procedures should migrate identically in all polyacrylamide gel electrophoretic systems. On the other hand, relatively minor differences in conditions encountered during isolation could cause such identical RNA molecules to assume different secondary structures that would cause them to migrate differently. It is also possible for RNA molecules of the same size but entirely different primary structure to migrate identically in a number of electrophoretic systems. In the case of PSTV and CEV, components of identical mobility are observed in 98% formamide-10% acrylamide gels (Figure 15E), while no such identical components

are seen in any of the other four sets of conditions selected for study here (Figure 15A-D). In light of the considerations discussed above, no conclusions about the possibility that PSTV and CEV are identical can be drawn from these comparative polyacrylamide gel analyses.

On the other hand, the results of electrophoretic studies of this type may contain clues to the structure of viroid RNA's. Simple RNA molecules such as E. coli 5S rRNA have been observed to form two or more discrete bands upon electrophoresis in 10% gels (Ikemura and Dahlberg, 1973). In addition, studies in which two components were resolved upon electrophoresis of formylated PSTV RNA have been reported (Diener and Smith, 1973). Thus, it is possible that viroid RNA exists in more than one native structure, or that there is more than one stable configuration which it can adopt once it is obtained in highly purified form. Because, in our studies, the two components sometimes occur in approximately equal amounts (e.g. Figure 15C) it seems unlikely that all of the second components observed could be due to minor contamination of viroid preparations by host RNA species. In any case, it should be possible to shed further light on the subjects of structure and composition of the infectious RNA's of viroids by combining polyacrylamide gel fractionation with two-dimensional RNA fingerprinting analyses.

Comparative Fingerprinting of PSTV and CEV RNA's: PSTV and CEV RNA's were labeled in vitro with ^{125}I. Although this procedure leads to some random breakdown of the RNA, both PSTV and CEV were shown to contain major components migrating at the position of the intact species in 10% polyacrylamide gels after iodination was carried out.

Fingerprints of RNase T1 digests of the iodinated viroid RNA's are shown in Figure 16. RNase T1 cleaves after all G residues, producing oligonucleotides such as CpGp, UpGp, CpUpGp etc. Only those fragments containing C are visible in autoradiograms of fingerprints since C is the only residue labeled to a significant extent in the chemical iodination procedure (see Chapter I). Since the small oligo-

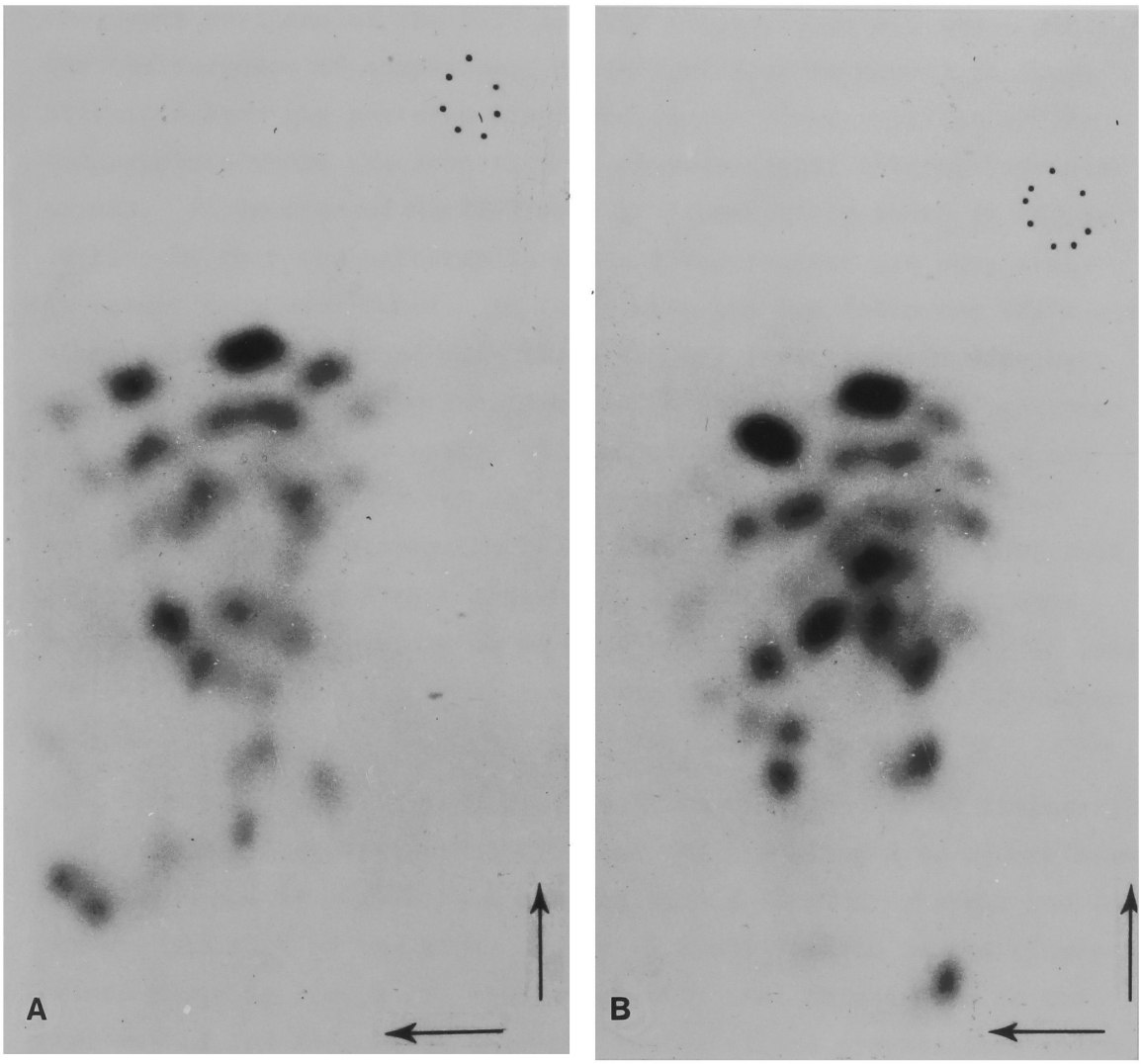


Figure 16

Ribonuclease T1 fingerprints of ^{125}I-labeled PSTV and CEV RNA. About 1 x 10^6 dpm of ^{125}I-labeled RNA were mixed with 10 micrograms of bacteriophage f2 RNA and digested with 2 micrograms of RNase T1 in 2 microliters of 0.01 M Tris.HCl (pH 7.5) - 1 mM EDTA for 40 minutes at 37° C. Reaction mixtures were fingerprinted as described in Chapter I. In the configuration shown here, the first dimension was from right to left and the second was from bottom to top as shown by the arrows. (A) PSTV RNA iodinated in vitro to a specific activity of 12 x 10^6 dpm per microgram. (B) CEV RNA iodinated in vitro to a specific activity of 18 x 10^6 dpm per microgram.

nucleotides have a high probability of occurring in any RNA molecule, the upper portions of the PSTV and CEV fingerprints are quite similar. For the purposes of comparison, it is therefore necessary to focus attention upon the patterns displayed in the lower portions of the fingerprints where the longer, more characteristic oligonucleotides appear. Inspection of the PSTV and CEV fingerprints shown in Figure 16 reveals that the patterns in these lower regions are completely different from each other. In fact, when the two iodinated RNA's were mixed together, digested with RNase T1, and fingerprinted together, all of the large oligonucleotides (14 of the total 25 to 27 oligonucleotides) with the exception of a minor one which migrates as the lowest leftward spot in the CEV fingerprint, could be identified separately. Since oligonucleotides which run in different positions always represent different sequences, then the minimum difference between the primary sequences of PSTV and CEV is 13.3% (i.e., at least one base change in each of 14 fragments of average length 7.5 nucleotides).

If the X-ray films used to detect the distribution of radioactivity in the T1 fingerprints of PSTV and CEV are allowed to become overexposed, a set of minor spots emerges with a complexity about two to three times that of the major family of spots visible in the fingerprints shown in Figure 16. It is probable that this is due to the presence of low amounts of host RNA in the viroid preparations which could represent as much as 10% of the total iodinated material. One alternative possibility is that these low intensity spots are the result of lowered iodination efficiency in regions of the RNA with secondary structure. This has been ruled out by carrying out test iodinations at temperatures up to 80° C--well above the melting point of the viroid RNA's (Diener, 1972; Semancik et al., 1973b). Identical fingerprints were obtained in these studies (unpublished results). A final possibility is that the infectious RNA species is not the major component of the viroid RNA preparations. If this is the case, then the major RNA species would at least have to be induced by viroid

infection since this component is not seen in 20% polyacrylamide gels of RNA from uninfected plants (Diener, 1972). Then it would seem a remarkable coincidence that the infectious RNA migrates in these gels exactly with a host RNA induced by its presence. One definitive way to check this possibility would be to determine the sequence of the major component of the viroid RNA preparation, synthesize RNA with this sequence, and then test its infectivity. Alternatively, continued study of the major components by determining, for example, whether or not their specific cleavage leads to loss of infectivity, may reveal whether or not they represent the infectious species.

Two-dimensional fingerprinting analyses of the pancreatic RNase digestion products of iodinated PSTV and CEV are shown in Figure 17. Since this enzyme cleaves after all pyrimidines, the average size of fragments produced is about half that produced by RNase T1 treatment. Each fingerprint has six spots corresponding to oligonucleotides larger than trimers, three of which have identical mobilities. The average size of these large pancreatic RNase-resistant oligonucleotides is 5.1 bases. Thus, the minimum difference between PSTV and CEV RNA sequences estimated from these fingerprints is 9.8% (i.e., at least one base change in three out of six fragments of average length 5.1 nucleotides).

Thus, we have shown by comparative two-dimensional fingerprinting analyses of ^{125}I-labeled RNA from PSTV and CEV that the major RNA components of these two viroids are not identical. (If, as discussed above, a minor component should turn out to be responsible for PSTV and CEV infectivity, then the formal possibility remains that these two agents are identical.) The minimum difference between the primary structures of the two major species is about 13% as estimated from a comparison of fingerprints of RNase T1 digests (Figure 16) and about 10% as estimated from fingerprints of pancreatic RNase digests (Figure 17). These estimates are based on the assumption that oligonucleotides which contain C are representative of the whole molecule. While the technique of comparative fingerprinting analysis is extremely sensitive for detecting minor differences in closely related RNA molecules (i.e.,

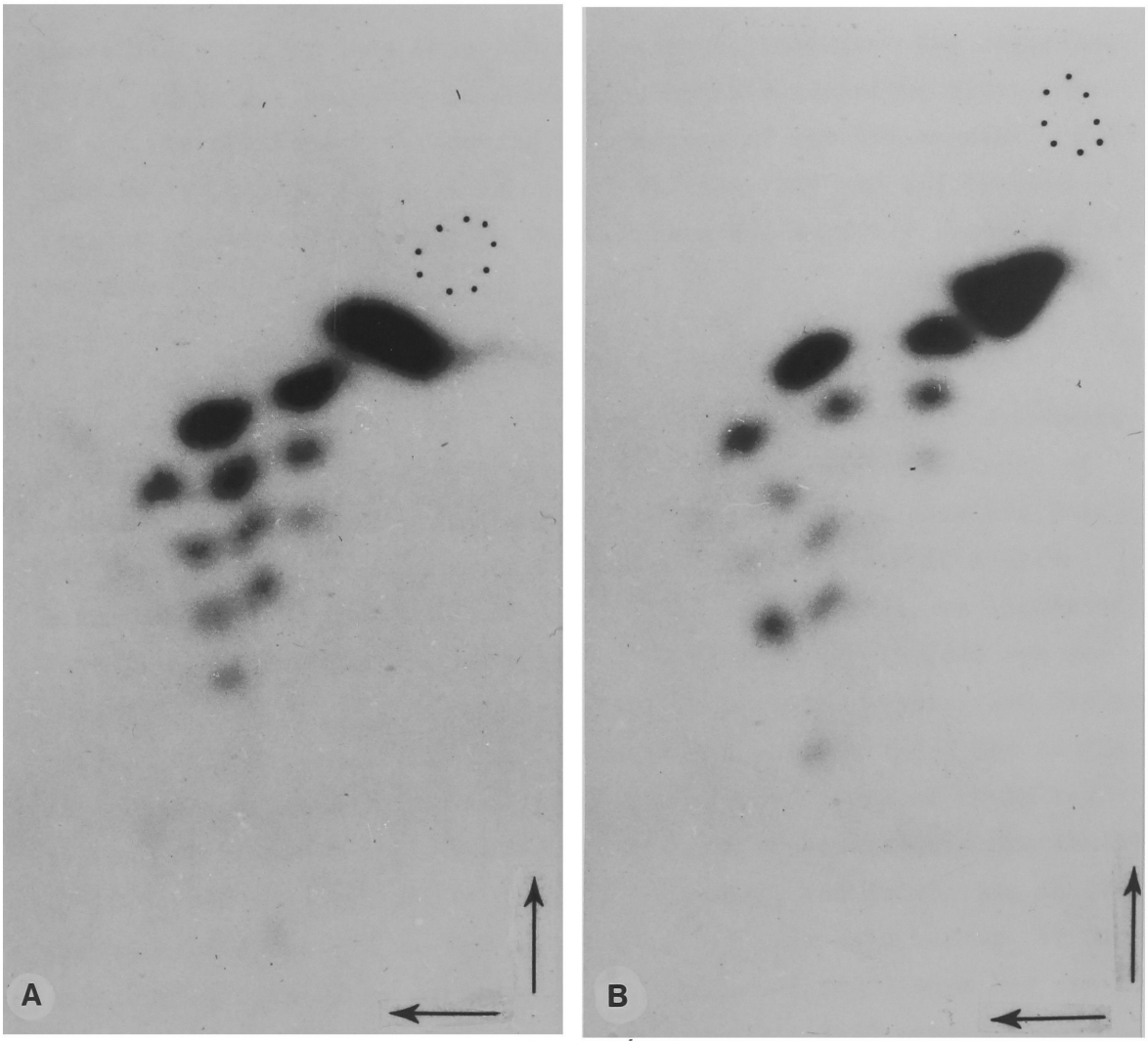


Figure 17

Pancreatic ribonuclease fingerprints of ^{125}I-labeled PSTV and CEV. About 1 x 10^6 dpm of ^{125}I-labeled RNA were mixed with 10 micrograms of bacteriophage f2 RNA and digested with 2 micrograms of pancreatic ribonuclease in 2 microliters of 0.01 M Tris.HCl (pH 7.4) - 1 mM EDTA for 30 minutes at 37° C. Two-dimensional fingerprinting analyses were carried out as detailed in Chapter I. The origin is at the lower right; the electrophoretic first dimension was from right to left, and the second dimension (homochromatography) was from the bottom of the picture to the top. (A) PSTV RNA iodinated in vitro to a specific activity of 12 x 10^6 dpm per microgram. (B) CEV RNA iodinated in vitro to a specific activity of 18 x 10^6 cpm per microgram.

those differing by less than 10% in sequence, Robertson and Jeppesen, 1972), it is not possible to determine, by this technique alone, what, if any, is the actual percentage relatedness of two RNA species. Further study will be required to reveal whether PSTV and CEV contain regions of partial homology or whether they are entirely unrelated in sequence.

A Search for Double-stranded RNA within PSTV and CEV

Because of the dramatic inhibition of in vitro protein synthesis by negligible amounts of double-stranded RNA and the involvement of double-stranded RNA in interferon induction, it seemed possible that the pathogenicity of viroid RNA might also be an effect of double-stranded RNA. To test for the presence of such regions, we incubated PSTV (1 microgram per reaction) and ^{125}I-labeled CEV (20,000 cpm per reaction) with E. coli Ribonuclease III (Robertson, Webster and Zinder, 1968). This enzyme is capable of reducing double-stranded RNA to TCA (trichloroacetic acid) solubility under the conditions of these reactions. Each viroid underwent four different treatments: incubation at 37° C for 40 minutes with no RNase III present, and incubation in the presence of RNase III in low salt buffer, or high salt buffer, or high salt buffer with added magnesium. This range of conditions has recently been shown to allow RNase III to demonstrate its full capacity to cleave RNA (see Chapter III; Dunn, 1976).

After incubation with RNase III, the viroid samples were run on a 7.5% polyacrylamide gel as shown and described in Figure 18. No change is observed in the amount of material in the viroid gel band for any of the reactions. Furthermore, when PSTV was incubated with RNase III and then tested for infectivity by Dr. T. O. Diener, no significant loss in infectivity was detected (Diener, personal communication, 1974). Since RNase III requires either an extended region of perfect double-stranded RNA (about 25 or more base pairs) or a highly specialized RNA sequence (Robertson and Hunter, 1975; Robertson and Dunn, 1975; see Chapter III) in order to cleave an RNA molecule, we conclude that neither PSTV nor CEV contain such a region.

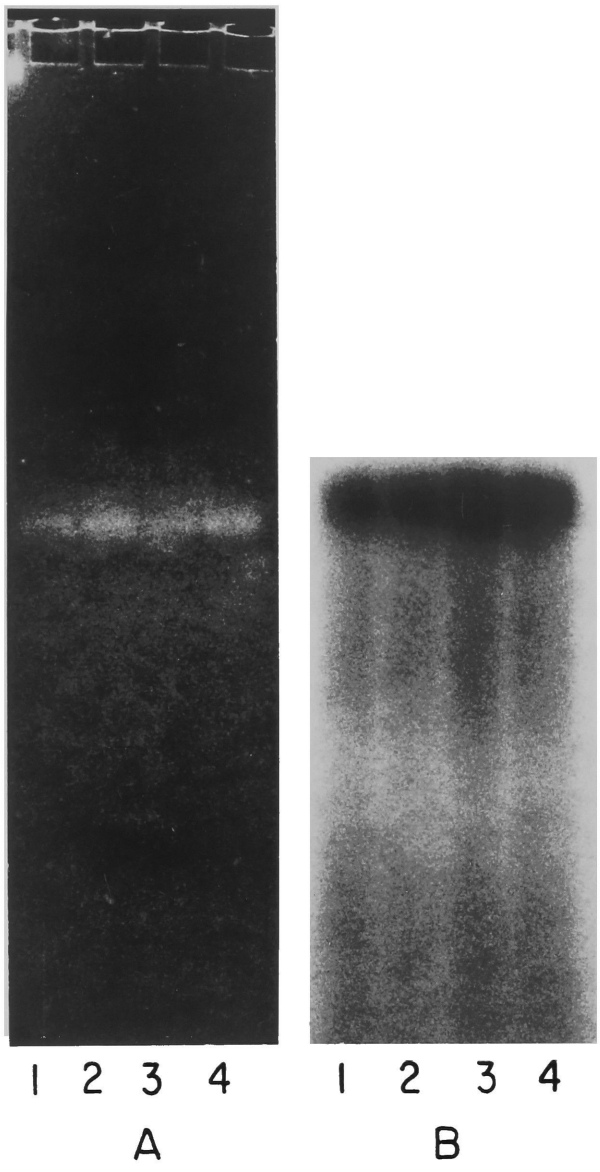


Figure 18

E. coli RNase III treatment of PSTV and CEV. 1 microgram of PSTV or 20,000 cpm of ^{125}I-labeled CEV were incubated at 37° C for 40 minutes either with no RNase III (lane 1), or with RNase III in high salt buffer (lane 2), in low salt buffer (lane 3), or in high salt buffer with magnesium ions (lane 4) under conditions that would render double-stranded RNA TCA-soluble in 20 minutes. (A) PSTV treated as described with RNase III, run on a 7.5% polyacrylamide gel, stained with ethidium bromide and photographed. (B) ^{125}I-labeled CEV treated in the same way as the PSTV (except here the position of the CEV was detected by X-ray film).

Characterization of Other Viroid RNA's

As outlined in the introduction to this chapter, when a viroid is discovered, it is important to fingerprint its RNA since the RNA sequence of a viroid represents its only known intrinsic property which is unique. The fingerprint will give a good indication of the degree of purity of the sample and show whether a unique species of RNA is present that is missing in the uninfected plant RNA.

Two such studies have already been carried out. In each case, RNA was purified in Beltsville by T. O. Diener and sent to us for fingerprinting assays. The samples were repurified by Whatman CF11 cellulose chromatography to remove trace contaminants that inhibit the iodination reaction (such as soluble acrylamide, Tris, SDS, etc.). After ethanol precipitation, one microgram was subjected to iodination and T1 and pancreatic RNase fingerprints performed. Such an analysis of CSV RNA is represented in Figure 19. Visual inspection reveals that this RNA is highly purified (free of significant contamination by host RNA), and has a complexity similar to that of PSTV and CEV. Figure 20 shows an identical study carried out on the RNA species that appears during cadang-cadang infection. In this case, gel bands were cut out from the healthy RNA gel fractionations from the position corresponding to the new RNA species observed in the RNA from diseased plants. When the T1 and pancreatic RNase fingerprints of these RNA's were compared (Figures 20 and 21) it was obvious that the new RNA species associated with diseased tissue is not present in RNA from healthy tissue in detectable amounts. Furthermore, this new RNA species has a complexity very much like the three viroids PSTV, CEV and CSV.

Progress on the Purification of Viroids

There are two basic technical problems to be overcome before it will be possible to isolate viroid RNA from a small sample of plant tissue (1-2 grams) in a form of high enough purity to allow iodination and fingerprint analysis. (i) A micro extraction procedure must be developed which can separate all of the RNA from other cellular com-

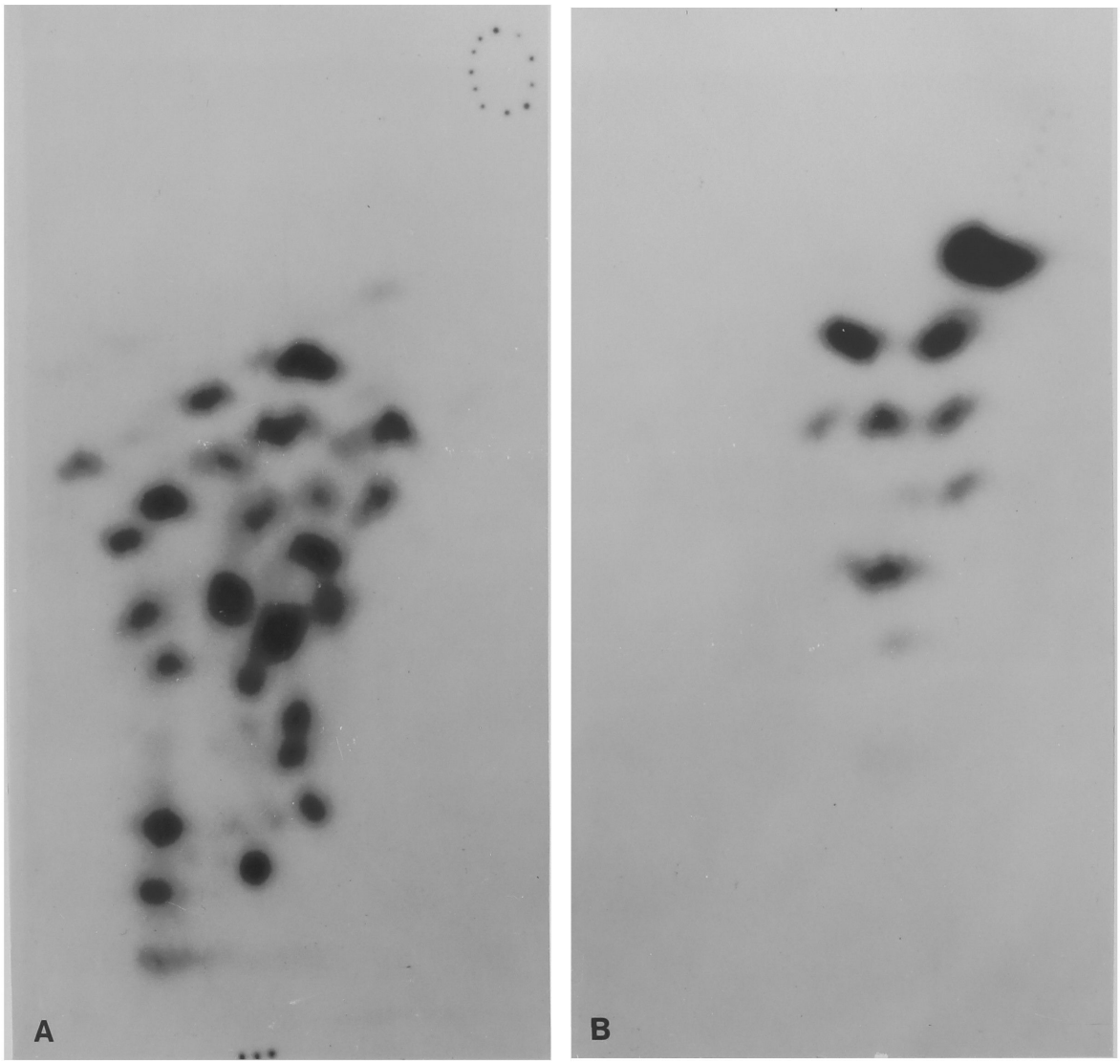


Figure 19

T1 and Pancreatic RNase fingerprints of chrysanthemum stunt viroid RNA (received for assay from T. O. Diener). Iodination and fingerprinting were carried out as described in Chapter I. (A) T1 fingerprint of CEV. (B) Pancreatic RNase fingerprint of CSV.

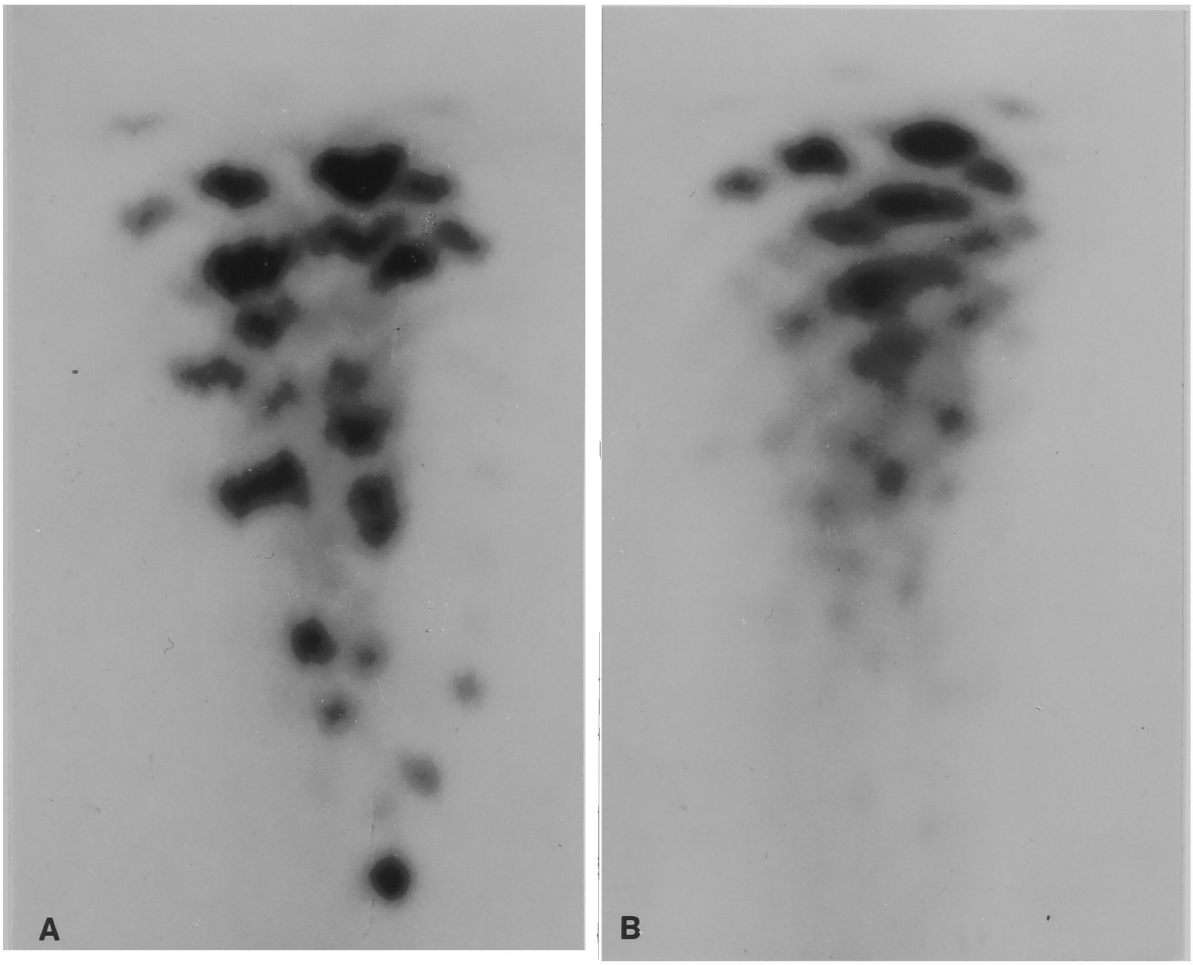


Figure 20

T1 fingerprints of iodinated cadang-cadang associated RNA species of coconut palms (A) and the RNA of similar mobility from healthy coconut palms (B).

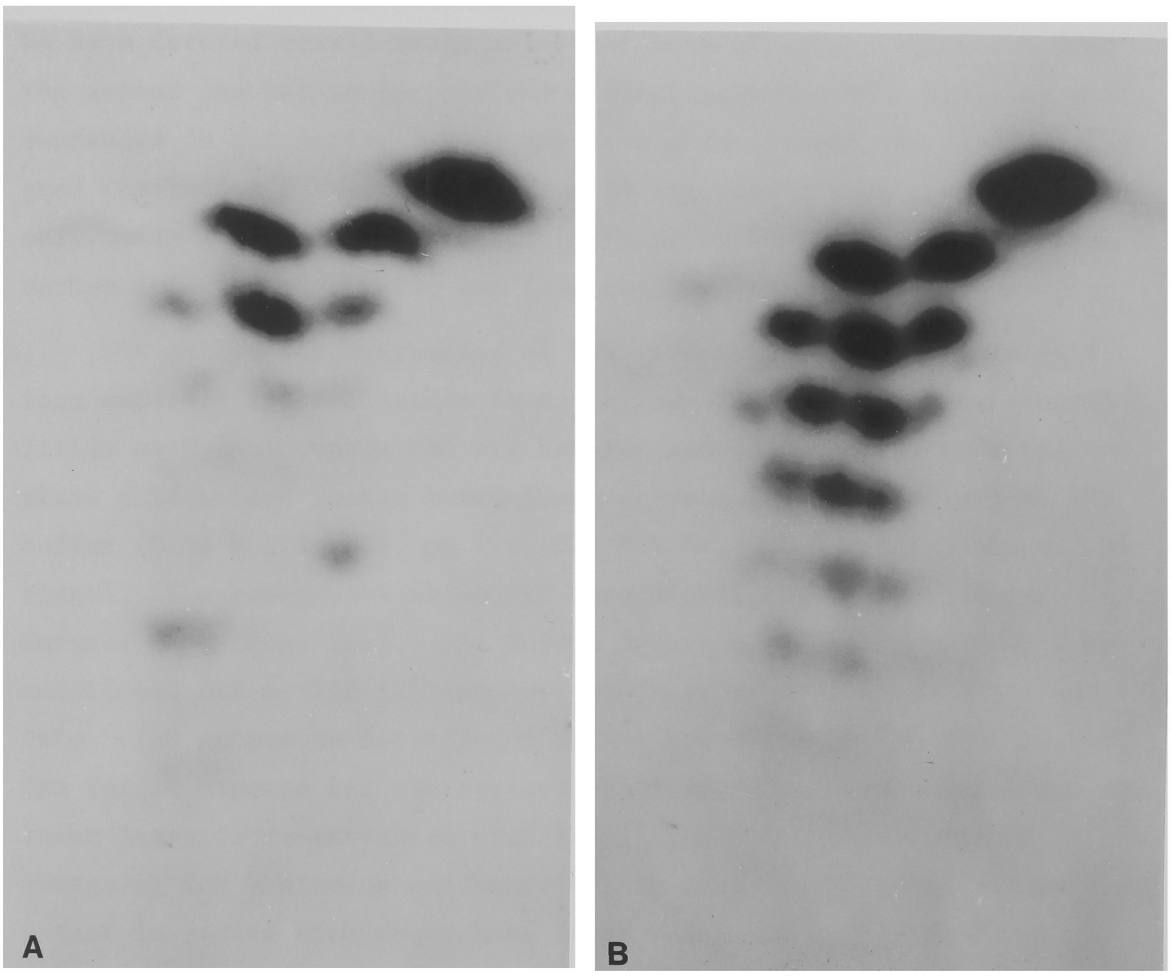


Figure 21

Pancreatic RNase fingerprints of iodinated cadang-cadang associated RNA species of coconut palms (A) and the RNA of similar mobility from healthy coconut palms (B).

ponents without RNA breakdown and without disproportionate loss of some RNA species. (ii) A method for extracting as little as 0.5 to 1 microgram of RNA from polyacrylamide gels in a carrier-free state must be developed which yields acrylamide-free RNA for subsequent iodination. We have devoted considerable effort to both of these problems. Only the second one has so far yielded a complete solution. While we have succeeded in extracting intact RNA from plant tissue and in obtaining good resolution between RNA species in the viroid size range in polyacrylamide gels, the relative proportions of individual RNA species varies in an unpredictable way from one preparation to another.

Our method for extraction of RNA from polyacrylamide gels in a form suitable for iodination is as follows. Gel bands containing as little as 0.5 microgram RNA are homogenized (25 strokes) in a teflon-glass siliconized Dounce homogenizer at room temperature in 1 ml TSE buffer (0.05 M Tris-HCl, pH 7.0, 0.1 M NaCl, 0.001 M EDTA) and 0.5 ml phenol. The homogenate is poured into a siliconized 12 ml Pyrex Sorvall centrifuge tube. The Dounce apparatus is rinsed out with an additional 0.5 ml TSE (10 strokes) which is added to the centrifuge tube. The sample is centrifuged in the Sorvall SS34 rotor at 15,000 rpm for 10 minutes and the aqueous upper layer removed carefully. The lower layer is re-extracted with 1 ml TSE and 0.2 ml chloroform by vortexing for 2 minutes and centrifuging again. The aqueous supernatant is pooled with the others in an unsiliconized 12 ml Pyrex Sorvall centrifuge tube. After 10 minutes of centrifugation at 15,000 rpm, the aqueous supernatant is removed and prepared for Whatman CF11 cellulose chromatography by the addition of an equal volume of ethanol. This mixture is loaded onto a CF11 column (prewashed with TSE and TSE-50% ethanol; column has bed volume of 0.2 cc and is packed in a 1 ml glass disposable syringe with glass wool plug), washed extensively with TSE-50% ethanol and then with TSE-35% ethanol. Finally the RNA is eluted from the column with 0.3 ml of TSE. HCl precipitation is then carried out as described in Materials and Methods.

After the second water wash following TEC resuspension of the precipitated RNA, the RNA is resuspended in 0.1 ml of water and its spectrum is analyzed using a Gilford Spectrophotometer equipped with microcuvettes. The samples are dried down for the last time and subjected to iodination.

Following iodination (which usually yields 1-5 x 10^6 cpm per microgram), the RNA is repurified on CF11 as described above. This time 10 micrograms of carrier RNA are added and the iodinated RNA is ethanol precipitated overnight (see Method and Materials, Chapter I). The RNA is then resuspended in 0.3 ml water and HCl precipitated. The HCl precipitate is resuspended in 30% TEC, dried down, resuspended in water, dried down, and washed twice more with water. Finally, the RNA is fingerprinted.

With this technology in hand, it is now possible to ask whether viroids "breed true" when grown in different host plants, or whether they are modified in some way. In order to investigate this question it is necessary to grow a particular viroid in a variety of different hosts, isolate the lower molecular weight RNA species, subject them to the purification scheme outlined above, iodinate and fingerprint them and finally compare the fingerprints with those studied previously. For example, we have inoculated Rutgers tomato plants with CEV originally isolated by J. S. Semancik from Gynura aurantiaca. There are three possible outcomes to such an experiment: the viroid fingerprint obtained from the infected tomato leaves could (i) be identical to CEV, (ii) be identical to PSTV, or (iii) have a new pattern. Either of the first two results would be unambiguous.

However, none of the RNA species characterized so far demonstrates a fingerprint like that of PSTV or CEV. Thus the third alternative--that the viroid RNA which replicates in tomato plants as a result of infection by CEV from Gynura aurantiaca is different from both PSTV and CEV--must be considered. For example, Figure 22 shows one of the viroid-sized RNA species extracted from CEV-infected tomato plants,

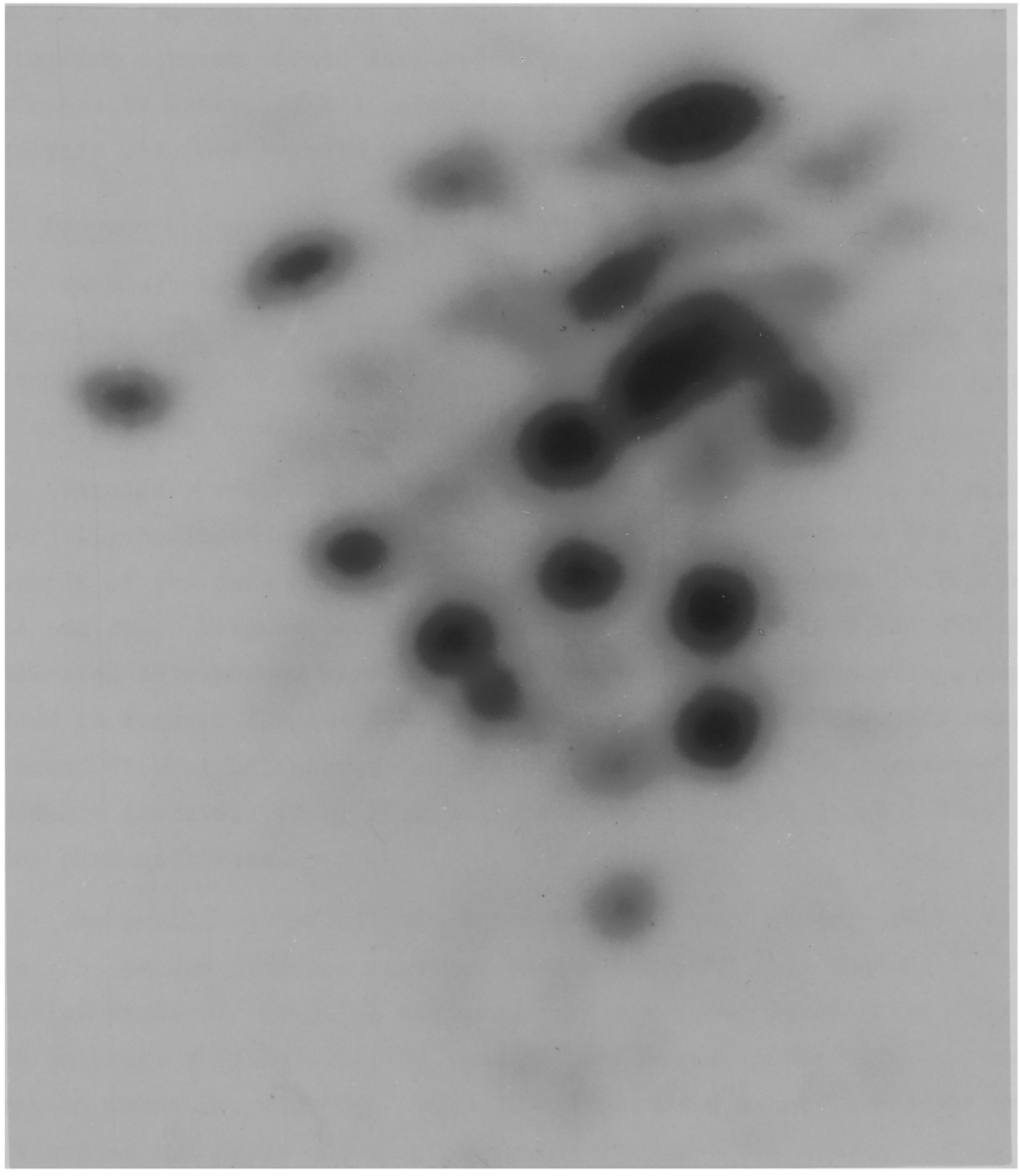


Figure 22

T1 fingerprint of a low molecular weight RNA isolated from a polyacrylamide gel band containing approximately one microgram of RNA (see text for method of extraction from gel). This RNA species was present in the low molecular weight RNA of CEV-infected tomato leaves.

and analysed as described above. Many other gel bands have been similarly treated. In the absence of a clear-cut result such as obtaining a known viroid fingerprint, it will be necessary to combine infectivity assays of all candidate RNA species with the fingerprint analysis outlined here in order to locate the new version of CEV.

Secondary Enzymatic Analysis of T1-Resistant Oligonucleotides

Each of the viroid RNA's fingerprinted so far has its own characteristic set of 10 to 12 T1-resistant oligonucleotides ranging in size from approximately 8 to 15 nucleotides in length that make up the unique pattern by which that viroid may be recognized.

Secondary enzymatic analyses of all such characteristic T1-resistant oligonucleotides have been carried out for each viroid RNA. An example of the data obtained in these studies is presented in Figure 24A and 24B. Oligonucleotides chosen for study from PSTV and CEV are indicated by numbered circles in the tracings of T1 fingerprints reproduced in Figures 23A and 23B. Aliquots of each oligonucleotide were treated with either pancreatic RNase or RNase U2 and the digestion products fractionated by high voltage electrophoresis at pH 1.9 on DE81 DEAE-cellulose.

The mobilities of all digestion products were calculated as a fraction of the distance migrated by the iodo-CMP (the fastest ^{125}I-labeled product). Careful comparison of the data obtained for PSTV and CEV indicate that none of the lage oligonucleotides are held in common between these two viroids. This fact raises our estimate for minimum difference and leads us to suspect that PSTV and CEV are not at all related. Nevertheless, it is of crucial importance of have on record thorough analyses of this kind. As mentioned at the end of the previous section, it may be necessary to screen a series of gel bands for an RNA species derived from a known RNA, but modified in some way by the host in which it is replicating. With the aid of complete secondary enzymatic analysis of all large T1-resistant oligonucleotides, it would be possible to detect potential subclasses of homology within such RNA's.

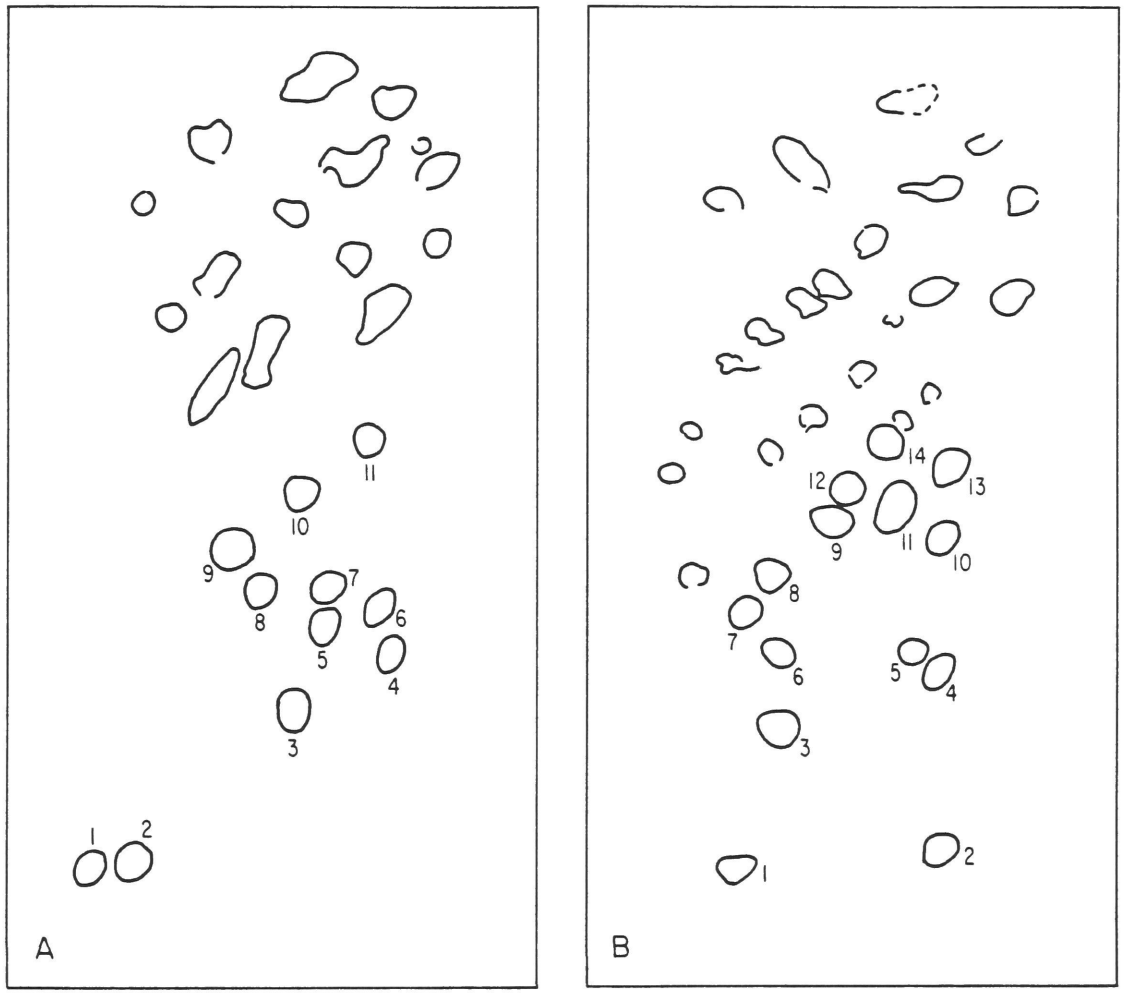


Figure 23

Tracings of T1 fingerprints of ^{125}I-labeled PSTV and CEV RNA. Each numbered circle corresponds to a T1-resistant oligonucleotide selected for secondary enzymatic analysis (see Figure 24) from either (A) PSTV or (B) CEV.

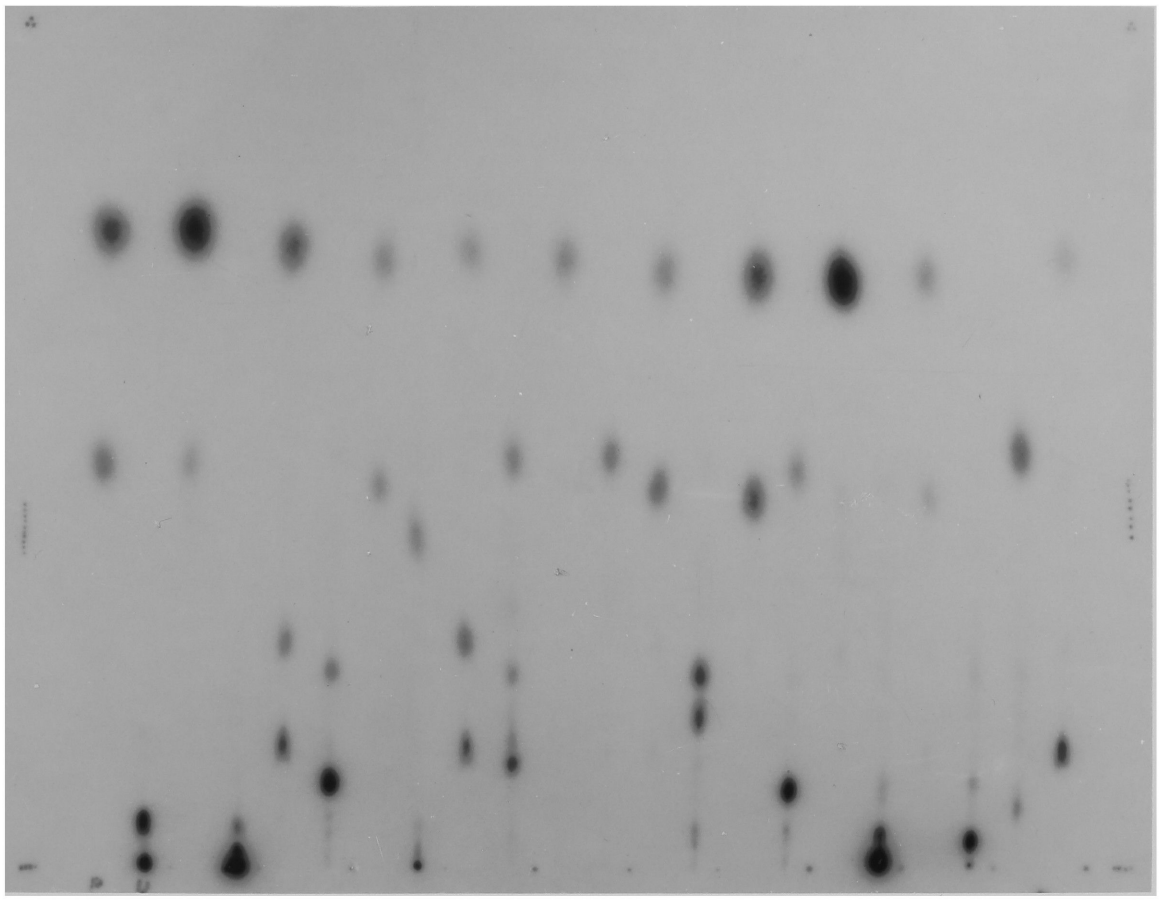


Figure 24A

Secondary enzymatic analysis of T1-resistant oligonucleotides from a T1 fingerprint of ^{125}I-labeled PSTV (see Figure 23A). T1-resistant oligonucleotides were eluted with 30% TEC, divided into two aliquots, and incubated with either pancreatic RNase or RNase U2 as detailed in Chapter I. Digests were spotted onto DE81 DEAE-paper and subjected to high voltage electrophoresis. Oligonucleotides #1 to #11 are arranged in numerical order from left to right with pancreatic RNase and RNase U2 digests for each oligonucleotide placed side by side (left to right, respectively).

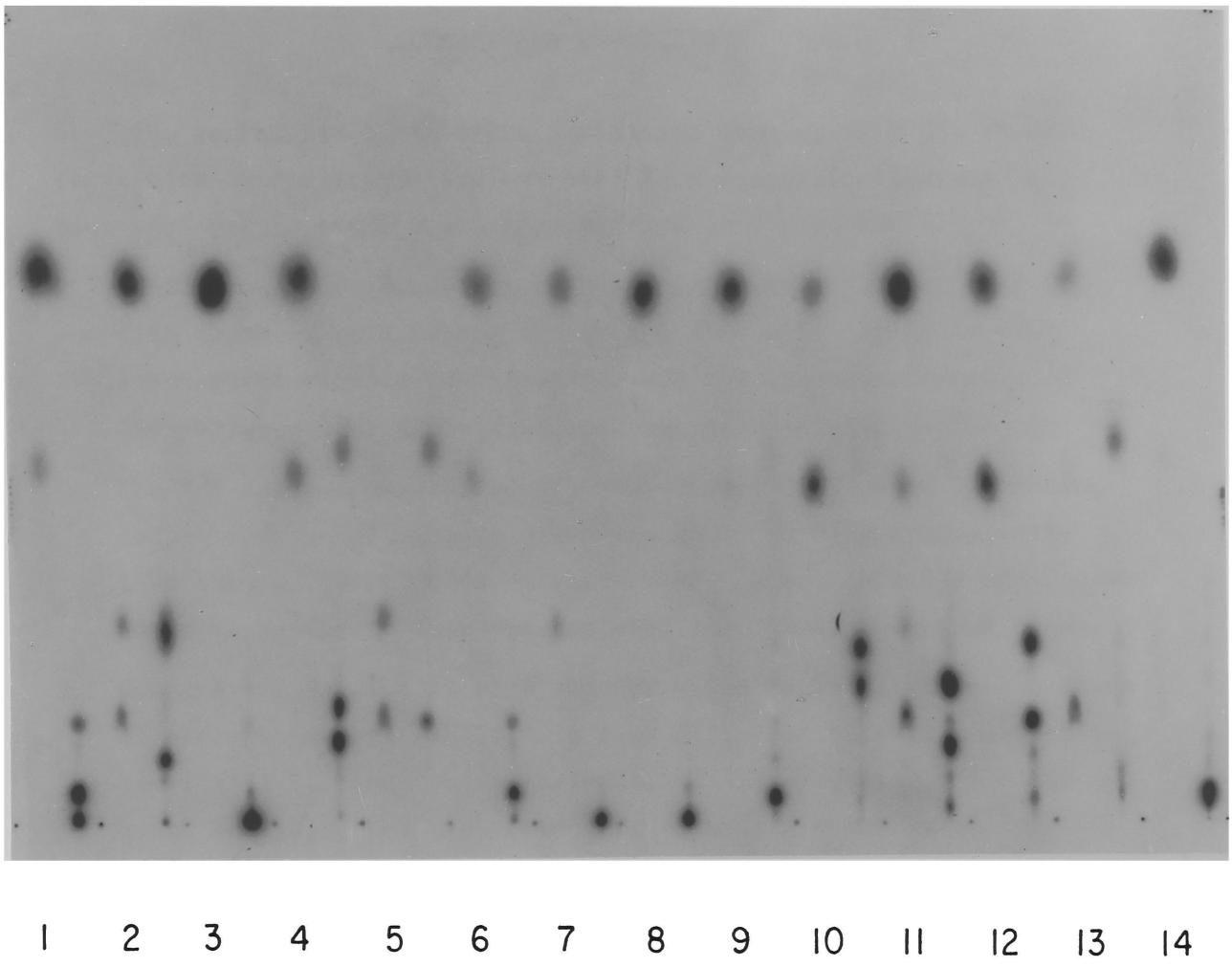


Figure 24B

Secondary enzymatic analysis of T1-resistant oligonucleotides from a T1 fingerprint of ^{125}I-labeled CEV (see Figure 23B). T1-resistant oligonucleotides were eluted with 30% TEC, divided into two aliquots, and incubated with either pancreatic RNase or RNase U2 as detailed in Chapter I. Digests were spotted onto DE81 DEAE-paper and subjected to high voltage electrophoresis. Oligonucleotides #1 to #14 are arranged in numerical order from left to right with pancreatic RNase and RNase U2 digests for each oligonucleotide placed side by side (left to right, respectively).

SUMMARY AND CONCLUSIONS

The techniques of in vitro iodination coupled with RNA characterization by fingerprinting and secondary enzymatic analyses have been employed to study several properties of viroid RNA's.

1. Comparison of T1 fingerprints of ^{125}I-labeled PSTV, CEV and CSV with those of HeLa 5S RNA and duck globin mRNA has shown that these three viroids contain major RNA species whose complexity lies between that of HeLa 5S and that of duck globin mRNA.

2. The RNA species associated with the diseased state in the cadang-cadang disease of coconut palms was shown to have a complexity like that of the viroids characterized. Also, this RNA species was shown to be absent from RNA isolated from healthy coconut palms.

3. Comparative studies of PSTV and CEV which included (i) measurement of electrophoretic mobility in five different polyacrylamide systems; (ii) T1 fingerprint analysis; (iii) pancreatic RNA fingerprint analysis; and (iv) secondary enzymatic analysis of the T1-resistant oligonucleotides led to the conclusion that these two viroids are not identical, as was previously thought to be the case, and may not even be related in RNA sequence.

4. Incubation of PSTV and CEV with E. coli Ribonuclease III neither affected infectivity nor affected mobility in 7.5% polyacrylamide gels under conditions that lead to total degradation of perfect double-stranded RNA. We conclude that there are no extensive regions (25 base pairs or longer) of perfect double-stranded RNA in either of these two viroids.

5. A new method for extracting low amounts of RNA (0.5 to 1.0 micrograms) from polyacrylamide gel bands in a highly purified form has been developed.

The new knowledge gained through these studies both about properties of viroids and about techniques for their study provides a solid foundation for future investigations of viroids and their interactions with plants.

CHAPTER III

POTENTIAL REGULATORY ROLES FOR RNA

Our original interest in the study of viroids arose through a consideration of potential roles for RNA in the control of gene expression. Because viroids are composed entirely of RNA (Diener, 1972), and do not appear to code for any protein product (Davies et al., 1974; Hall et al., 1974), it is possible that they interact directly with the regulatory machinery of the host cells and could constitute aberrant forms of regulatory substances.

In this chapter we will first consider the features of RNA which make it especially suitable for a regulatory role, then suggest a possible source for the proposed regulatory RNA, and finally describe possible modes of action. The work presented here is based in part on a paper by H. D. Robertson and E. Dickson (1974) and a paper by E. Dickson and H. D. Robertson (1976, in press).

Properties of RNA Which Make It Especially Suitable As a Regulatory Molecule

Molecules which have a primary role in the control of gene expression must be able to interact with the DNA in a highly specific manner. Proteins such as the *lac* repressor can select and bind with high affinity to a unique region within the DNA of *E. coli* and prevent transcription. However, the amount of information required to specify this interaction is about 1000 nucleotides since this regulatory protein is about 350 amino acids long (Gilbert and Müller-Hill, 1970). When the notion of repressor proteins is extended to eukaryotic systems, it becomes apparent that a second gene would be required to control transcription of the repressor protein gene, and this second gene action must also be regulated, and so on. In contrast, consider the efficiency of using RNA to select specific regions from within the DNA. RNA is particularly well suited to this task since it is capable of

binding to a single-stranded DNA by complementary base pairing. In addition, a sequence of nucleotides just 17 bases long is sufficient to specify a unique region within the DNA of the human genome (Britten and Roberts, 1969).

While RNA has a coding capacity equivalent to that of DNA, it is subtly different from DNA in that it contains a hydroxyl group on the second carbon of the ribose ring. Because of this chemical difference the organism can use enzymes which digest RNA rapidly while leaving DNA intact. Then abrupt changes in state of differentiation could be accomplished by rapid destruction of one set of RNA molecules, and creation of a new array.

Recent X-ray crystallographic studies (Ladner et al., 1975) of yeast phenylalanine tRNA at a resolution of 2.5 Å have shown that the presence of the 2'-hydroxyl group in RNA confers additional properties upon these molecules. This hydroxyl group can receive and/or donate protons in hydrogen bonding to various parts of the backbone of the polynucleotide or to the pyrimidine or purine rings. As a result, the tRNA folds up in regions in a complex way more reminiscent of protein structure than nucleic acid structure. Such "globular" regions are ideal for interaction with proteins. In other regions the tRNA demonstrates base pairing and helicity more conventionally associated with nucleic acid structure.

Because RNA can turn over rapidly, has a high coding capacity and can interact with high specificity with both DNA and protein, it is a prime candidate for the job of gene regulation.

A Source for Regulatory RNA in Eukaryotes

No function has yet been ascribed to the major class of RNA transcribed in eukaryotic nuclei. This heterogeneous nuclear RNA (hnRNA), so-called because it ranges in size up to 100 S and greater in some organisms, remains mainly in the nucleus where the bulk of it is broken down rapidly with only about ten percent ever reaching the cytoplasm (Attardi et al., 1966; Warner et al., 1966; Darnell, 1968). Some

authors have suggested that the rapid turnover in the nucleus of the majority of this RNA represents degradation of unnecessary transcripts (Scherrer and Marcaud, 1968; Brandhorst and McConkey, 1974; Williams and Penman, 1975). We find much more attractive suggestions such as that of Judd and Young (1973) that some of the "extra" RNA within huge transcripts could have a regulatory role within the nucleus where it could act to coordinate action of related genes. The conventionally held view of what happens to the extra RNA in the nucleus is that a battery of powerful nucleases attack and render it soluble as a result of rapid random cleavage. If, however, a function is assigned to this class of RNA, it is necessary to postulate the presence of enzymes within the nucleus which can release the regulatory RNA from the huge RNA precursors in a precise and reproducible way.

Evidence for the Presence of Processing Enzymes Within Eukaryotic Nuclei

RNase P, a well-characterized enzyme from E. coli, is a good prototype for the kind of enzyme to look for in eukaryotic nuclei (Robertson et al., 1972). In E. coli, tyrosine tRNA is released from a precursor RNA by cleavage of a single phosphodiester bond by RNase P as shown in Figure 25 (Altman and Smith, 1971; Robertson et al., 1972; Altman et al., 1974). The RNase P reaction is extremely sensitive to mutational changes in the sequence of the tRNA. For example, mutations at position 15 or 25 of the E. coli tyrosine tRNA drastically slow down the rate of cleavage of the precursor by RNase P (Altman and Smith, 1971; Altman and Robertson, 1973). In vivo in these mutants, the tRNA precursor is degraded by other enzymes before the RNase P cleavage can occur so that the mature mutant tRNA appears in much lower than wild-type amounts. On the other hand, it has been shown that when this mutant tRNA is released from its precursor by RNase P in vitro, it has a stability comparable to wild-type tRNA's (Altman and Robertson, 1973; Altman et al., 1974; Altman, 1975). Two important characteristics of this enzyme need to be stressed: (1) the signal

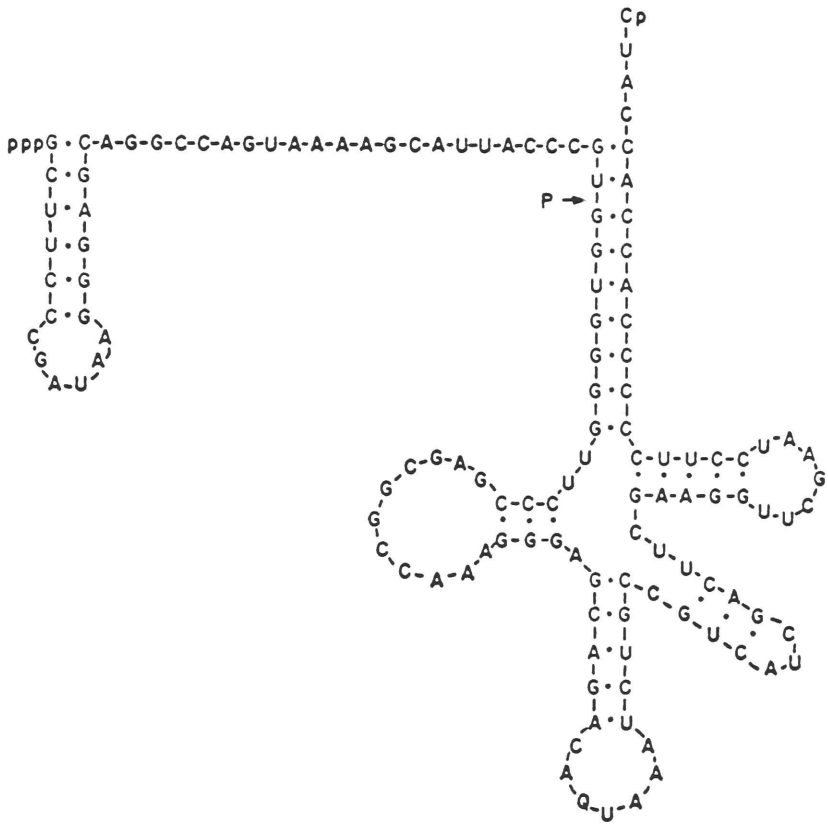
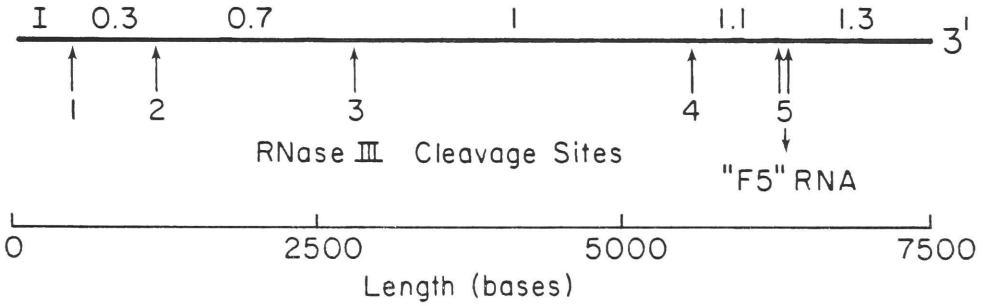


Figure 25

Sequence and possible secondary structure of E. coli $tRNA^{Tyr}$ precursor RNA (Altman and Smith, 1971) showing the site of RNase P cleavage 41 nucleotides from the 5' end of the precursor, at the site of the 5' terminus of the mature tRNA molecule (Robertson et al., 1972).

which specifies cleavage includes features of sequence and structure within a large region surrounding the point of cleavage and (2) proper cleavage releases a mature RNA species which is stable in the presence of the same nucleases that are capable of demolishing this same RNA region when it is part of the precursor molecule. A schematic representation of these features appears in Figure 26. This example serves to illustrate the concept of generating, by cleavage, a stable RNA molecule from a previously unstable RNA species.

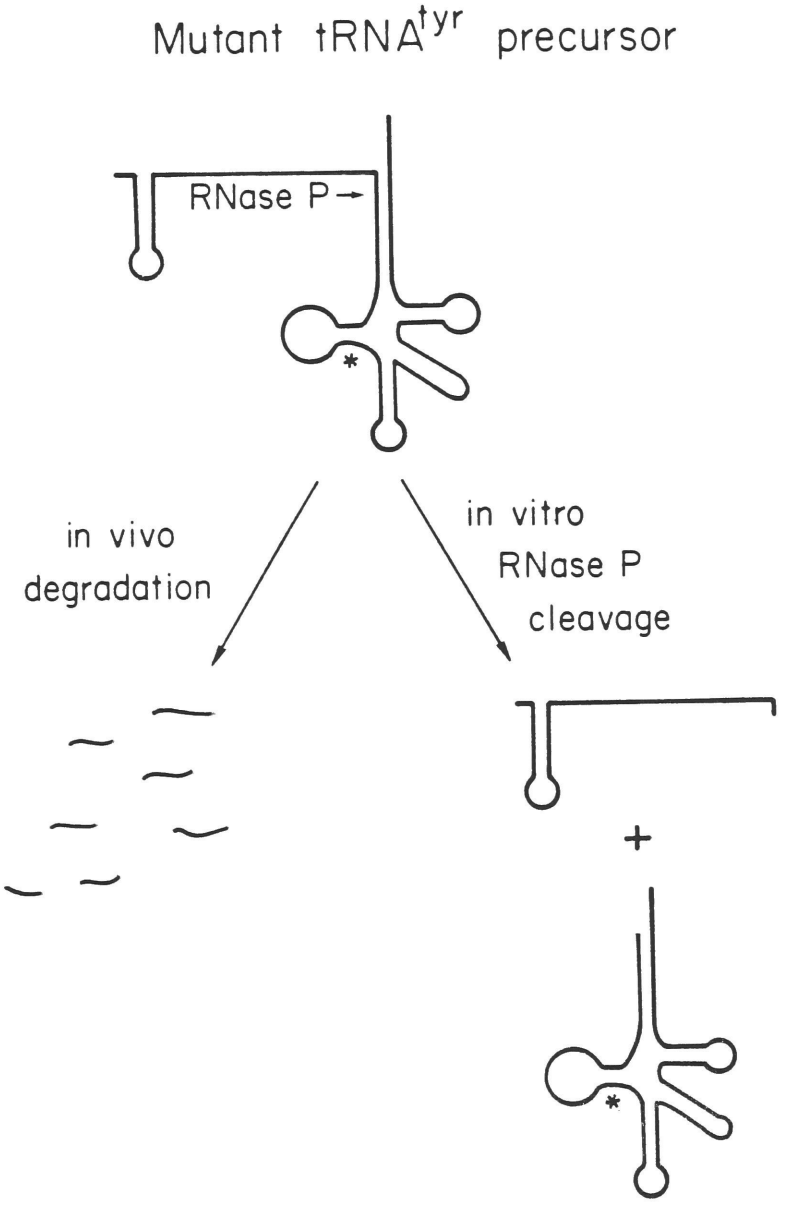
A second enzyme from E. coli--Ribonuclease III (Robertson et al., 1967; Robertson et al., 1968)--has recently been shown to be an extremely subtle RNA processing enzyme. When presented with perfect double-stranded RNA, RNase III yields a limit digest consisting of fragments about 15 nucleotides long (Robertson et al., 1968; Robertson and Dunn, 1975). On the other hand, bacteriophage T7 early mRNA precursor is cleaved both in vivo and in vitro by RNase III at exactly five sites resulting in the release of the five T7 early mRNA species from the precursor molecule as shown schematically below (Dunn and Studier, 1973a, 1973b).



The conjectures that specific cleavage sites in such single-stranded RNA may have double-stranded character (Dunn and Studier, 1973b; Robertson and Dunn, 1975) or may contain specific sequences (Robertson and Dickson, 1974; Rosenberg et al., 1974) can be tested directly by carrying out sequence analysis of RNA regions surrounding cleavage sites. Recent collaborative studies with H. D. Robertson and

Figure 26

Effects of a point mutation in the E. coli tyrosine tRNA sequence upon the metabolism of the tRNA precursor molecule. At the top of this figure is represented schematically the E. coli tyrosine tRNA precursor sequence (see Figure 25). This molecule can undergo a number of mutations, such as the one symbolized by (*) in the figure. This asterisk represents a base change from G to A at the position 25 bases from the RNase P cleavage site, which is the 5' terminus of the mature tRNA molecule. The effect of this mutation is to cause the near disappearance of the mature tyrosine tRNA from the cell, and accumulation of abnormally high levels of uncleaved tyrosine tRNA precursor (Altman and Smith, 1971). As shown in the lower left portion of the figure this mutant precursor molecule turns over in vivo, undergoing degradation without accumulating significant amounts of mature tRNA (Altman and Smith, 1971; Altman, 1975). The explanation for these events in vivo comes from studies on the in vitro cleavage of mutant tyrosine tRNA precursor by purified RNase P (Robertson et al., 1972; Altman et al., 1974). As shown in the lower right of the figure, proper cleavage occurs giving rise to the expected two fragments--the tRNA itself and the 41-nucleotide 5' terminal "extra" fragment. However, studies by Altman and his colleagues (Altman et al., 1974) have shown that the rate of cleavage by RNase P on mutant as compared to wild-type precursors is drastically lowered. Therefore, although the cleavage can occur with fidelity, the rate in vivo must be insufficient to prevent the mutant precursor from entering an alternative pathway leading to its complete breakdown. One of the most interesting features of this alternative pathway for precursor breakdown is that only the mutant precursor molecules enter it. Those few mutant tRNA molecules which do mature in vivo or in vitro are just as stable as measured by half life or stability in cell-free extracts as their wild-type counterparts (Altman and Robertson, 1973). Thus the vast majority of mutant precursor RNA's are destroyed in the cell even though they contain a potentially stable tRNA sequence within them.


Mutant tRNA^{tyr} precursor
RNase P
*
in vivo
degradation
in vitro
RNase P
cleavage
+
*

J. J. Dunn (Brookhaven National Laboratory) have allowed us to propose a sequence 29 bases long for the RNA fragment released by the double cleavage event that separates gene 1.1 from gene 1.3. This RNA species is referred to as the F5 RNA. In vitro transcription of T7 DNA utilizing ^{32}P-labeled ribonucleoside triphosphates, followed by RNase III digestion and polyacrylamide gel fractionation (Dunn, 1976 in press), yielded highly purified RNA suitable for sequence determination. Two-dimensional fingerprinting and subsequent extensive enzymatic analyses of the oligonucleotides (see Chapter I and Pieczenik et al., 1974) provided us with data (presented in Table IV-A & B) which allowed us to propose a sequence for this RNA species.

Figure 27 shows the nucleotide sequence of F5 RNA drawn with bars above and below to indicate the information yielded by each pancreatic RNase-resistant and RNase T1-resistant oligonucleotide analyzed. The overlapping information derived from these two sets of specific digestion products allowed us to arrange all of the fragments into a single molecule with no ambiguity.

F5 RNA is redrawn in Figure 28 to demonstrate the potential secondary structure. Work by others (personal communication from R. Kramer, M. Rosenberg and J. A. Steitz of Yale University) on sequence analysis of another RNase III cleavage site from this same mRNA precursor shows a great deal of homology of sequence. It therefore appears likely that RNase III recognizes elements of sequence as well as secondary structure. If secondary structure with stability of that shown in Figure 28 were sufficient by itself to constitute the signal for cleavage by RNase III, then species such as the RNA phage Band 23 from R17 or f2 (also shown in Figure 28) should also act as signals for such cleavage. However, RNase III does not cleave f2 or R17 RNA within Band 23 (Robertson et al., 1968; Robertson and Hunter, 1975) nor is isolated Band 23 a substrate for RNase III. Therefore, the signal within the RNA that specifies cleavage of T7 F5 RNA by RNase III must be more subtle than the mere presence of a short helical region. Rosenberg et al. (1974) have demonstrated the presence of the

Table IV-A

Analyses of RNase T1-resistant oligonucleotides from the RNA fragment separating bacteriophage T7 genes 1.1 and 1.3 (F5 RNA Species)

Number	Proposed Sequence	NTP[a]	Composition[b]	Pancreatic RNase products[c]	RNase U2 Products[c]
1	CUCUCUAG[G][d,e]	U	C*	C*	CUCUCUA*[e]
		C	U*	U*	CUCUCUA
		A	U*	U*	CUCUCUA*
		G	A*,G*	AG*[G]	CUCUCUA*,G*[G]
2	CCUUAG[U][e]	U	C*,U*,G*	C*,U,AG*[U]	CCUUA*,G*[U]
		C	C*	C*	CCUUA*
		A	U*	U*	CCUUA*
		G	A*	AG*	CCUUA*[G]
3	pUAAG[G][f]	U	pUp*[f]	pUp*	pUA*
		C	--	--	--
		A	pUp*,A*	pUp*,AAG*	pUA*
		G	A*,G*	AAG*[G]	A*,G*
4	UCG[C]	U	--	--	--
		C	U*,G*	U*,G*	UCG*[C]
		A	--	--	--
		G	C*	C*	UCG*
5	AG[U]	U	G*	AG*[U]	G*
		C	--	--	--
		A	--	--	--
		G	A*	AG*	A*[G]
6	UG[G]	U	--	--	--
		C	--	--	--
		A	--	--	--
		G	U*,G*	U*,G*	UG*[G]

Table IV-A (cont.)

Number	Proposed Sequence	NTP[a]	Composition[b]	Pancreatic RNase Products[c]	RNase U2 Products[c]
7	G[U/A/C]	U	$\overset{*}{G}$	$\overset{*}{G}$	$\overset{*}{G}$
		C	$\overset{*}{G}$	$\overset{*}{G}$	$\overset{*}{G}$
		A	$\overset{*}{G}$	$\overset{*}{G}$	$\overset{*}{G}$
		G	--	--	--

Table IV-B

Analyses of pancreatic RNase-resistant oligonucleotides from the RNA fragment separating bacteriophage T7 genes 1.1 and 1.3 (F5 RNA Species)

Number	Proposed Sequence	NTP[a]	Composition[b]	RNase T1 Products[c]
1	AGGAGU[G]	U	$\overset{*}{G}$	$A\overset{*}{G}$[U]
		C	--	--
		A	$\overset{*}{G}$	$\overset{*}{G}$[A]
		G	$\overset{*}{A},\overset{*}{G}$	$\overset{**}{AG}$[G], $\overset{*}{A}G$
2	AAGGU[C]	U	$\overset{*}{G}$	$\overset{*}{G}$[U]
		C	$\overset{*}{U}$	$\overset{*}{U}$[C]
		A	$\overset{*}{A}$	$\overset{*}{A}AG$
		G	$\overset{*}{A},\overset{*}{G}$	$AA\overset{**}{G}$[G]
3	GGC[C]	U	--	--
		C	$\overset{*}{G},\overset{*}{C}$	$\overset{*}{G}$[C]$\overset{*}{C}$[C]
		A	--	--
		G	$\overset{*}{G}$	$\overset{*}{G}$[G]
4	AGU_{OH}[g]	U	$\overset{*}{G}$[g]	$A\overset{*}{G}$[U]
		C	--[g]	--
		A	--[g]	--
		G	$\overset{*}{A}$[g]	$\overset{*}{A}G$

Table IV-B (cont.)

Number	Proposed Sequence	NTP[a]	Composition[b]	RNase T1 Products[c]
5	GC[U]	U	$\overset{*}{C}$	$\overset{*}{C}$[U]
		C	$\overset{*}{G}$	$\overset{*}{G}$[C]
		A	--	--
		G	--	--
6	pU[A]	U	$\overset{*}{p}$Up	$\overset{*}{p}$Up
		C	--	--
		A	pU$\overset{*}{p}$	pU$\overset{*}{p}$
		G	--	--
7	C[U/G]	U	$\overset{*}{C}$	$\overset{*}{C}$
		C	--	--
		A	--	--
		G	$\overset{*}{C}$	$\overset{*}{C}$
8	U[C/A/U]	U	$\overset{*}{U}$	$\overset{*}{U}$
		C	$\overset{*}{U}$	$\overset{*}{U}$
		A	$\overset{*}{U}$	$\overset{*}{U}$
		G	--	--

Footnotes

[a]NTP refers to the ribonucleoside [alpha-^{32}P] triphosphate precursor used to make the oligonucleotides analyzed in a given line.

[b]Base compositions were determined as described in Chapter I. A dash in a column signifies that the particular oligonucleotide was not present in digests of the F5 RNA species.

[c]Proposed oligonucleotide compositions (other than those with only one base) listed in these columns were assigned both by mobility of the digestion products on unidimensional electrophoresis at pH 1.9 or 3.5 with reference to known standards, and by analysis of the base composition of these further digestion products. Asterisks denote those bases found to be labeled after base composition analysis.

[d]Square brackets indicate the proposed nearest-neighbor base.

[e]Completion of these proposed sequences required further analysis both of base composition of the F5 RNA fragment, and of the products of CMCT-blocked pancreatic RNase digestion.

[f]The presence of 5' terminal pUp was also confirmed by two-dimensional electrophoretic analysis (Robertson and Dunn, 1975).

[g]Identity of this oligonucleotide with the 3' terminus of the fragment is based on the failure to recover a pyrimidine labeled by any of the 4 NTP's. Thus the U-residue which labels AG must be located at the 3' end of the oligonucleotide lacking a 3' phosphate. This is to be expected since RNase III cleavage yields a 5' phosphate and a 3' hydroxyl (Robertson and Dunn, 1975). This sequence (AGU_{OH}) can be distinguished from AG by its mobility on the two-dimensional fingerprinting system.

sequence pGAU at the 5' termini of several cleavage products of the bacteriophage T7 early mRNA precursor by RNase III, thereby supporting the view that sequence recognition could be involved in the cleavage signal. It is probable that sequence analysis of at least several more RNase III cleavage sites will have to be carried out before a coherent view of the minimum and sufficient specifications for such a site can be developed.

In this context, it is useful to consider the various levels of complexity an RNA molecule can demand in interactions with other biological substances. Figure 29 illustrates the sort of sequence and structural features which could, formally speaking, comprise a signal within single-stranded RNA. It would appear, from sequence data now available, that the only two processing enzymes which have been studied in detail so far, RNase P and RNase III from E. coli, both require signals of the sort represented in Figure 29 (c), (e), or (f). This sort of signal would allow the existence of enzymes of the type which cleave a variety of different RNA molecules whose sequences bear little, if any, resemblance to each other.

This description of the processing reactions of two bacterial RNases has had two purposes: (i) to demonstrate the type of specific cleavage potentially available in all living systems, and (ii) to illustrate features within RNA molecules that may be involved in recognition by these enzymes. While no eukaryotic enzymes have been so extensively characterized as these bacterial examples, there is considerable evidence for their existence, particularly within cell nuclei.

An enzyme with specificity identical to that of RNase P has recently been described in the KB strains of human tissue culture cells (Altman et al., 1974; Bothwell and Altman, 1975). A second specific enzymatic activity, RNase NU, has been found both in mammalian cells and in E. coli (Bothwell and Altman, 1975). Thus, it is apparant that, not only do eukaryotic cells contain RNA processing enzymes, but also

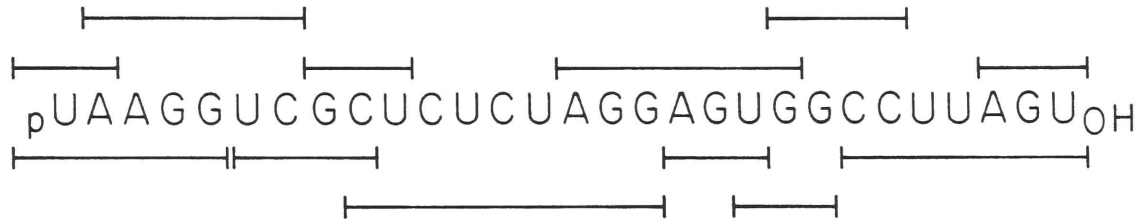


Figure 27

Nucleotide sequence of bacteriophage T7 F5 RNA. The 5' terminus of the RNA fragment is at the left side of the picture. The bars placed above the sequence indicate the pancreatic RNase-resistant oligonucleotides plus an additional base, in each case at the 3' terminus, whose presence was determined by nearest neighbor analysis. Bars beneath the sequence indicate RNase T1-resistant oligonucleotides plus the nearest neighbor at their 3' termini (see Tables IV-A and IV-B).

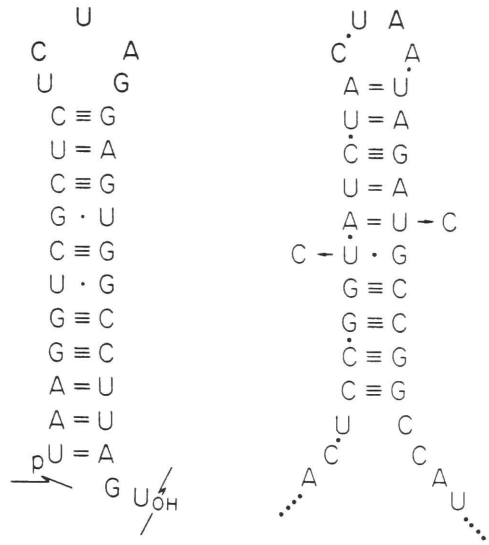


Figure 28

Proposed sequence and structure of an RNase III cleavage site from bacteriophage T7 early mRNA precursor and a similar uncleaved region from R17 bacteriophage RNA. On the left is a schematic representation of the sequence and structure proposed for the 29-base fragment (F5 RNA) released by RNase III cleavage of bacteriophage T7 early mRNA precursor between genes 1.1 and 1.3. The two RNase III cleavage points are located at the two termini of the fragment.

On the right side of the drawing is a schematic representation of Band 23 from the RNA of bacteriophages R17 and f2. The sequence represented in the hairpin loop is that of R17 (Nichols, 1970), while the two indicated C substitutions are those found in the analogous regions of f2 RNA (Nichols and Robertson, 1971). This sequence represents the termination point of the RNA phage coat cistron, and the UAA triplet at the top of the hairpin, together with the UAG immediately following, are the actual terminators for this cistron in vivo. Thus the left side of the hairpin codes for this carboxy terminus of the protein, and the right side represents intercistronic space prior to the beginning of the replicase cistron. It is noteworthy that three of the four bases at the top of the hairpin, as well as ten of the twenty bases in the stem, are exactly homologous with the T7 sequence shown at the left. Since both of these RNA regions consist of intercistronic sequences, this degree of homology in sequence and structure may reflect present or past identities of function. However, despite its similarity to the T7 hairpin, the Band 23 region of phage RNA is apparently not an RNase III cleavage site (Robertson and Hunter, 1975; Robertson et al., 1968), a fortunate outcome for the survival of the RNA phage which would otherwise undergo lethal cleavage of its genome.

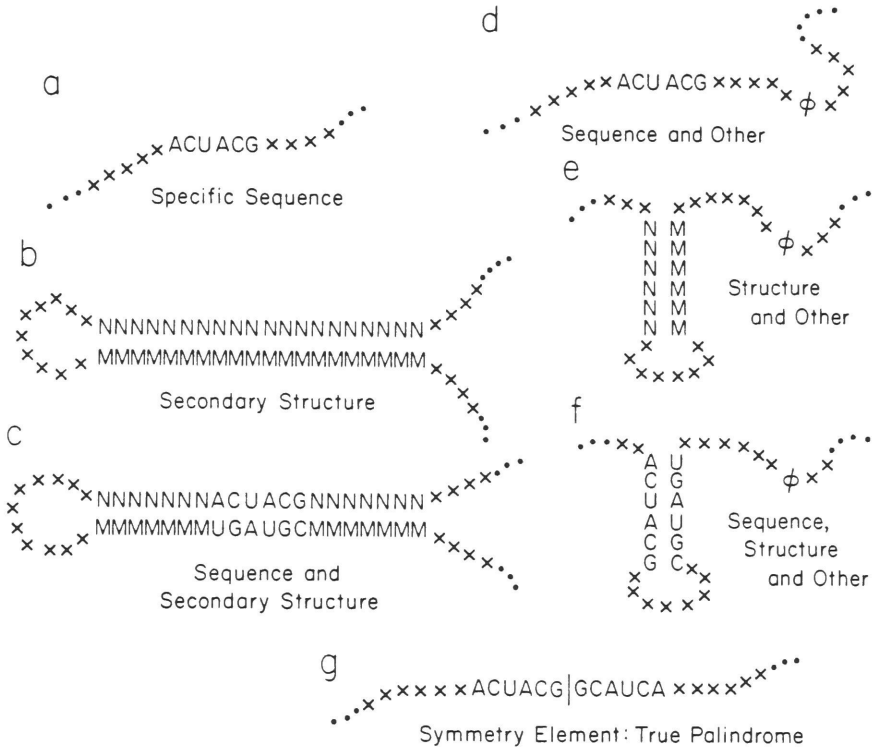


Figure 29

Potential signals within single-stranded RNA molecules for cleavage by specific RNA processing enzymes. The most straightforward such signal would be a specific base sequence as shown in (a); alternatively, the RNA could fold back to give a hairpin loop with a perfectly base-paired stem (b). Either such feature could provide a cleavage site. In addition, as shown in (c), these two features could be combined so that a specific sequence is recognized within a double helical region. A more complex sort of recognition would arise if processing enzymes recognized features at a distance from the site of cleavage. Such "other" features might include tertiary structures like the tRNA cloverleaf, or simply a second region of specific sequence or structure. These "other" sites could be combined with specific sequence or helical secondary structure to provide the cleavage sites shown in (d) - (f). Finally, symmetry elements such as the true palindrome (Pieczenik et al., 1974) illustrated in (g) could be recognized in any of the above ways.

these activities may be conserved in evolution (Altman and Robertson, 1973; Robertson and Dickson, 1974).

Activities which cleave RNA:RNA duplexes have also been detected in ascites tumor cells (Robertson and Mathews, 1973) and in the nuclei of human KB and HeLa cells (Altman et al., 1974; Schlessinger et al., 1974). The heterogeneous nuclear RNA in mammalian cells is an example of an RNA class in higher organisms which appears to require processing. In order to see if this processing might involve double helical regions, H. D. Robertson and W. Jelinek carried out the following experiment. ^{3}H-labeled HeLa cell hnRNA was fractionated by sucrose density gradient centrifugation and the RNA sedimenting in the 70S region pooled. When this material was incubated with E. coli RNase III essentially all of the RNA shifted from 70S to 14-18S without detectable creation of smaller fragments. This result was not entirely unexpected in light of previous findings that three percent of HeLa cell hnRNA can be isolated as stable double-stranded fragments (Jelinek et al., 1973; Jelinek and Darnell, 1972; Ryskov et al., 1973). These regions (isolated by resistance to pancreatic RNase) are themselves substrates for RNase III cleavage (Robertson and Dickson, 1974). Thus it appeared likely that RNase III cleavage occurs in vitro at such double-stranded sites within intact hnRNA molecules. Thus these dsRNA regions are distributed fairly evenly along most hnRNA molecules, and could serve as potential specific cleavage sites in vivo. This expectation is made more concrete by knowledge of the fact that the dsRNA regions of HeLa hnRNA turn over 100% in the nucleus, without any substantial amount surviving for transport to the cytoplasm.

Inspired by the observation that considerable homology exists between RNase III processing sites in the T7 early mRNA and by other indications that E. coli RNA processing could be sequence- as well as structure-dependent, we decided to investigate whether the dsRNA regions of HeLa hnRNA might also contain common sequence elements. Such a finding might be expected if the HeLa RNA regions indeed repre-

sent sites for specific cleavage analogous to their E. coli counterparts. As shown in the RNase T1 fingerprints in Figure 30, the isolation of dsRNA from hnRNA results in a disproportionate loss of complexity. The grey background which is the earmark of complexity gives way to simplicity. Six characteristic oligonucleotides which dominate the lower portion of the dsRNA fingerprint are also visible (especially numbers 1, 2 and 3) in the fingerprint of unfractionated hnRNA, emphasizing the fact that these oligonucleotides are greatly enriched within the unfractionated hnRNA. H. D. Robertson (personal communication, 1975) has shown that RNase III treatment of dsRNA followed by T1 fingerprinting analysis results in a dramatic reduction in the intensity of spots 1, 2 and 3, indicating that these oligonucleotides are located wholly within double-stranded regions.

Sequence analysis of the six major T1-resistant fragments of dsRNA has revealed that each represents a unique sequence, that their total information content is about 50-60 nucleotides (probably 10% of the sequence complexity of the dsRNA), and that there is potential for forming complementary regions between the sequences of some of the fragments (Robertson, Jelinek and Dickson, 1976, manuscript in preparation). One of our purposes in carrying out the RNA sequencing project was to create a well-characterized probe with which to search for RNA processing enzymes in mammalian nuclei.

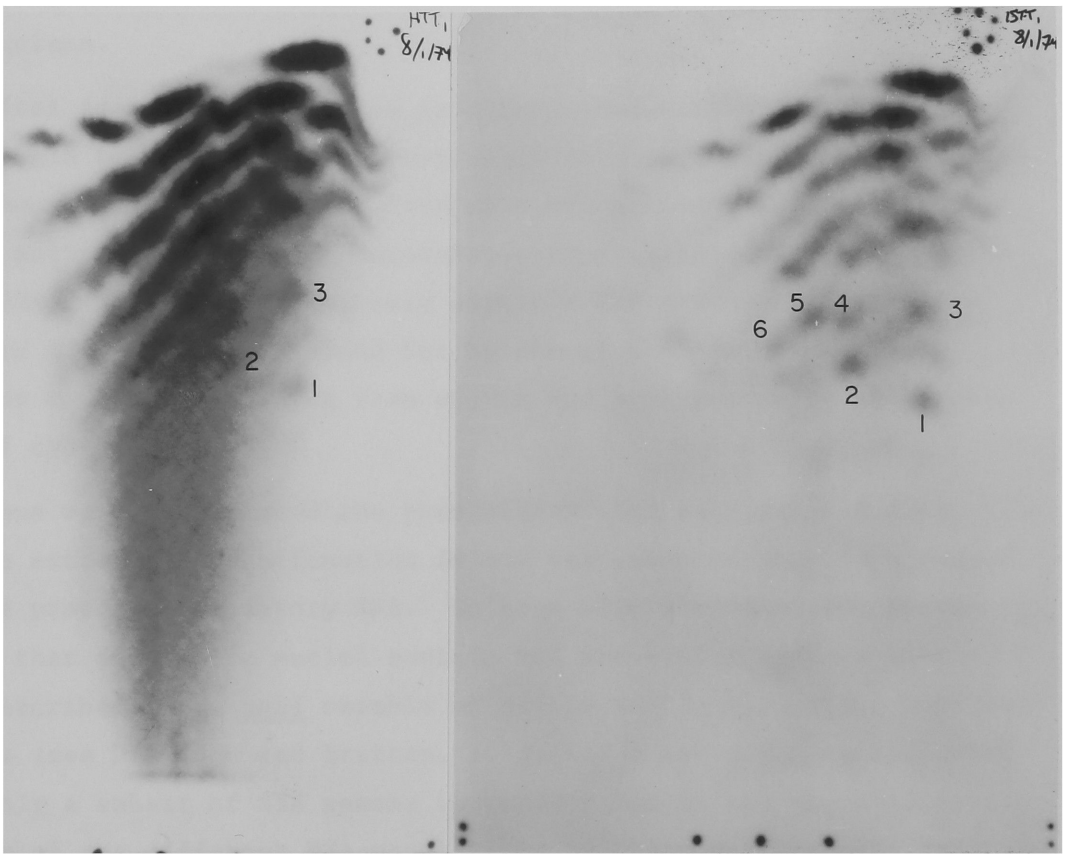
The presence of RNA processing enzymes within mammalian nuclei can be inferred from recent studies of Adenovirus RNA metabolism (Philipson et al., 1974; Sharp et al., 1974). Transcripts from both strands of the Adenovirus genome are present in the host nucleus both during early and during late periods of the infection cycle. During the early time, however, only the early mRNA reaches the cytoplasm as mature mRNA, while regions containing late gene transcripts turn over rapidly in the host nucleus.

Thus, at the time of Adenovirus infection, the host nucleus must contain specific RNA processing enzymes which can recognize the Adeno-

Figure 30

RNase T1 fingerprints of HeLa heterogeneous nuclear RNA and double-stranded regions of HeLa hnRNA. ^{32}P-labeled RNA from HeLa cell nuclei was prepared as described elsewhere (Robertson, Jelinek and Dickson, 1976, manuscript in preparation), and 90% of the resulting hnRNA was converted to RNase-resistant dsRNA regions. Equal aliquots of total hnRNA and the dsRNA regions (about 5 x 10^6 cpm) were digested with two microliters of RNase T1 solution (1 mg/ml) for 40 minutes and then subjected to conventional two-dimensional fractionation as described in Chapter I using thin-layer homochromatography on DEAE cellulose as the second dimension.

In the left panel, it is evident that unfractionated hnRNA yields an RNase T1 fingerprint of great complexity with little resolution of individual spots of a size larger than 4-5 nucleotides (those nearest the top of the figure). In the right panel, in contrast, a very simple pattern is observed, with no spot visible of a size greater than 9-10 nucleotides. Comparison of this right-hand panel with fingerprints of RNA molecules of known sequence (tRNA's; HeLa 5S RNA; viroid RNA's; globin mRNA) leads us to estimate that this pattern represents about 500 bases worth of information. The six prominent RNase T1-resistant oligonucleotides marked 1-6 have been prepared repeatedly and their nucleotide sequences largely determined. Each represents a unique sequence and their total information content is about 50 bases. Furthermore, closer inspection of the left-hand panel reveals detectable spots in the positions marked 1, 2 and 3. When these oligonucleotides were eluted and characterized in comparison with their counterparts from the fingerprint shown in the right-hand panel, we can show that they are indeed the same sequences. Therefore, the unfractionated bulk hnRNA is enriched for these characteristic sequences to an extent sufficient for them to be detected without going through the initial preparation of dsRNA. We can estimate that there must be an enrichment of these sequences in the RNA population of at least several hundred fold in order for these sequences to be detectable in this manner.


HTT 1
8/1/74
3
2
1
ISAT 1
8/1/74
6
5
4
3
2
1

virus early mRNA regions within the longer transcripts and cleave them out, thereby permitting their maturation. Since this cleavage occurs before production of any Adenovirus proteins is possible, the RNA processing enzyme(s) responsible must pre-exist and have a role in the host nucleus.

After Adenovirus infection is established, transcripts of late genes begin to appear in the cytoplasm. An interesting possibility is that the switch from early to late mRNA production is accomplished by the action of a new RNA processing enzyme capable of releasing the late mRNA's from their previously unstable RNA precursor. This new activity could either be coded for by one of the Adenovirus early genes or could be turned on from within the host genome by the presence of the Adenovirus.

Thus we have explored the possibility that eukaryotic nuclei contain extra RNA whose function is not yet known to provide a source for the proposed regulatory RNA. We have also discussed the possibility that eukaryotic nuclei contain RNA processing enzymes of the type described in E. coli capable of highly subtle cleavage. Numerous studies (see Davidson and Britten, 1973, for a review) have indicated that only a subset of the genome is transcribed in any one cell type. If each of the different RNA populations in these various cell types each undergoes specific RNA processing of the type described above, then each will yield a characteristic set of RNA fragments reflecting the state of differentiation of that particular cell type. By their existence, these products of specific RNA metabolism constitute a set of signals indicating the state of differentiation. We would like to explore the possibility that the cell utilizes these specific signals to amplify gene coordination.

THE ROLE OF RNA FRAGMENTS IN MODELS OF GENE CONTROL

Previously Proposed Roles for Regulatory RNA

(1) Specific RNA's could act as inducers, interacting with repressor-like proteins bound to specific DNA sequences, releasing the proteins and allowing transcription (Jacob and Monod, 1969). Conversely, in a manner analogous to the action of cyclic AMP in the action of CAP factor upon lac operon transcription, RNA could be used to confer specificity upon proteins which select sites of transcription (Beckwith and Zipser, 1970; Bekhor et al., 1969; Huang and Huang, 1969; Zubay et al., 1970).

(2) Britten and Davidson have suggested that an RNA could bind to a complementary region in the DNA, turning on the adjacent gene in an unspecified manner (Britten and Davidson, 1969). For example, RNA could displace tightly bound proteins to permit transcription (Nichols and Robertson, 1971). RNA could also interact directly with the DNA, thus changing its structure in a way that alters its availability to polymerase (see Paul, 1972; Sutton, 1972).

(3) Controlling RNA's might be a special class of mRNA which turns over rapidly and whose product is a regulatory protein (Davidson and Britten, 1973). This idea has a number of features in common with those which primarily invoke protein molecules for the control of transcription (Georgiev, 1969; Gierer, 1973; Scherrer and Marcaud, 1968; Tsanev and Sendov, 1971). Since there have been a number of detailed considerations of the idea that regulatory proteins could regulate part or all of gene expression in eukaryotes (Georgiev, 1969; Gierer, 1973; Scherrer and Marcaud, 1968; Tsanev and Sendov, 1971), we have concentrated here upon the additional compatible possibilities made manifest by considering the products of RNA processing in gene regulation.

One common feature of these three suggested modes of action for RNA is that they are all catalytic in nature. Thus a single molecule turns on the transcription of a genetic unit which then stays on

until a second independent event occurs to turn it off. A second feature is that in most cases the RNA species which carries out the proposed role has not been identified, although several authors suggest that a rapidly labeled fraction of hnRNA contains the appropriate species (Holmes et al., 1974; Judd and Young, 1973; Monahan and Hall, 1974). One exception is the case of chromosomal RNA studied by Bonner and his colleagues (Bekhor et al., 1969; Holmes et al., 1974; Huang and Huang, 1969), a heterogeneous collection of small RNA molecules which can be isolated from chromatin. This RNA appears to confer a specific pattern of transcription in vitro using DNA or chromatin as template. Recent hybridization competition experiments (Holmes et al., 1974) have indicated that chromosomal RNA and hnRNA have some homologous regions, supporting the suggestion that this chromosomal RNA is a cleavage product of hnRNA.

A third feature of the above models is that they do not take the probable presence of specific RNA processing enzymes fully into account. In light of recent experimental evidence, it seems likely that reproducible cleavage of primary transcripts yields specific RNA's which could then be invoked in considering control events.

A Novel Role for RNA Fragments

RNA fragments could influence gene expression by serving as primers for the transcription of additional RNA molecules. The 3' terminal region of an RNA fragment could locate a complementary region of the DNA, and the primer would then be extended by an RNA polymerase using the complementary strand DNA as template. Thus, the RNA transcript would include the primer fragment covalently linked to the new RNA molecule. It has been proposed that a mechanism of this sort is important in the DNA replication of both prokaryotes (Sugino et al., 1972; Watson, 1972) and eukaryotes (Verma et al., 1971). For example, Dahlberg et al. (1974) have demonstrated that a $tRNA^{trp}$ synthesized in uninfected chick cells is the primer which is found

covalently linked to Rous Sarcoma Virus DNA during the first round of reverse transcriptase action.

To explore the possibility that RNA priming could occur in transcription, it is necessary to study the properties of the RNA polymerases involved. E. coli RNA polymerase will utilize an RNA oligonucleotide as primer under certain conditions in vitro, in preference to ATP or GTP. Using synthetic DNA's containing sequences from the E. coli $tRNA^{tyr}$ gene, Kleppe and Khorana (1972) showed that there is a three-fold preference for a complementary primer hexanucleotide over ATP or GTP when RNA synthesis was carried out using a 29-base pair double-stranded DNA template. When the appropriate single DNA strand was utilized (Terao et al., 1972), an even greater preference for the primer over mononucleoside triphosphates was observed. No studies on the ability of any eukaryotic RNA polymerase to initiate synthesis using RNA primers complementary to the DNA template have so far been reported.

To explore the idea that some of the specific cleavage products of extra regions of RNA precursors could play a regulatory role by acting as primer for RNA synthesis, we have made three assumptions: (i) mRNA is transcribed as part of a longer RNA molecule; (ii) these precursors are cleaved in the nucleus by highly specific processing enzymes to yield mature mRNA and extra fragments; and (iii) RNA primers can initiate transcription in vivo.

Figure 31 shows a way by which the processing of a particular mRNA precursor could generate primers which cause the coordinated transcription of other genetic units. In the middle of the figure is a representation of part of a genome indicating boundaries of three genetic units. These are all different from the gene from which the mRNA precursor in the upper part of the figure was transcribed. Regions of the extra portion of the "pre-mRNA" molecule which can be cleaved out by specific processing events and subsequently used as primers are indicated. We postulate that these short RNA's could locate complementary regions of DNA within other genetic

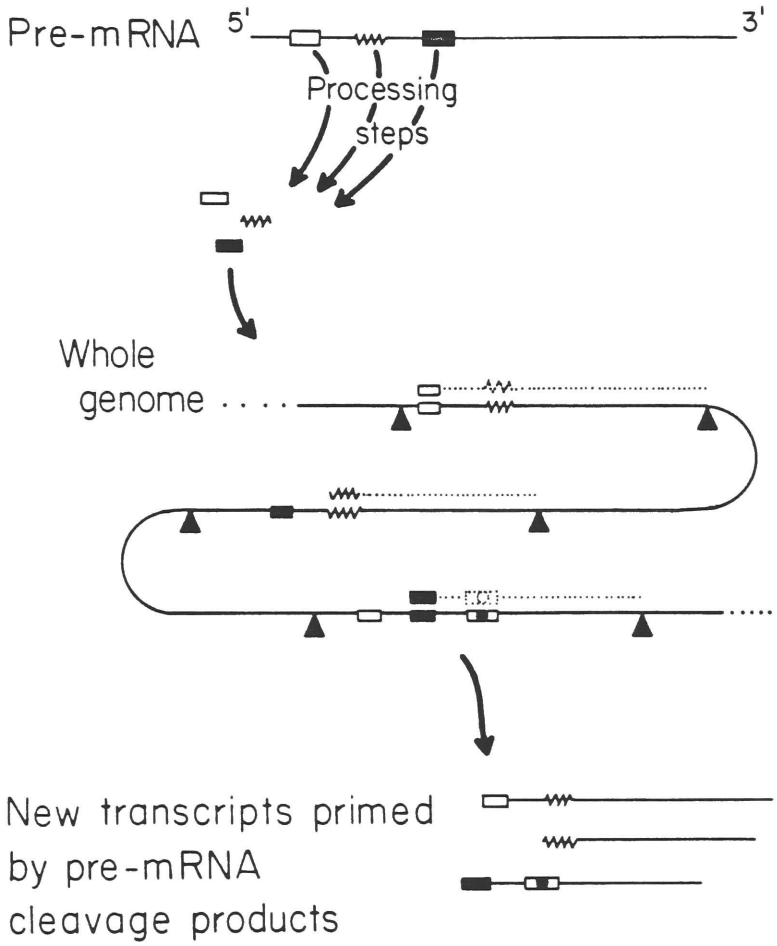


Figure 31

A possible role for RNA primers in the coordination of gene expression. The symbols (▭), (∿∿) and (▬) represent special regions within the RNA which can be cleaved from an mRNA precursor by highly specific RNA processing enzymes and subsequently used to prime RNA transcription. These same symbols are also used to depict sites in the DNA which encode these special RNA's. Three such RNA fragments resulting from the processing of the mRNA precursor molecule shown at the top of the drawing are illustrated here priming transcription at three other sites within the DNA. Thus, the RNA primer starts out as a cleavage product from within one mRNA precursor and subsequently becomes covalently attached to the 5' end of another one. Depending on the particular RNA primer selected, the mRNA precursor transcribed would contain different numbers of regions which could give rise to primers upon processing.

units and prime transcription, serving as the 5' ends of the new mRNA precursor molecules as shown at the bottom of the figure. Note that all of the genes depicted in the figure have multiple start points of this sort. Depending on the requirement for coordination with other genes, genetic units could contain a few or very many such regions which could both encode and receive primer signals. Thus by uniting the idea of specific processing of RNA precursors with the possibility that transcription of some genes might require priming by specific RNA's we have arrived at a novel means by which, without invoking protein synthesis and elaborate feedback controls, we could account for the control of some genes merely by the expression of others. Once having formulated this basic outlook, it is possible to consider elaborations on this theme. Nevertheless, the central idea--that RNA regions linked to a given mRNA precursor can be specifically and reproducibly removed and used to cause directly the synthesis of another--is one of compelling simplicity which merits careful consideration.

Consequences of Primed Transcription Mediated by Specific RNA Processing

Stoichiometry. One distinctive feature of any such mechanism of control which depends upon priming to generate new RNA transcripts is that it is stoichiometric (here defined as a process in which a specific control element functions and is consumed in the event which it regulates). As shown in Figure 31, the single mRNA precursor shown at the top of the figure releases only one copy of each of the three potential primers depicted. When these function at other points on the genome, each can prime only one transcription event. For example, the open rectangle could prime at its complementary site in either the upper or lower genetic unit shown containing its complement, but not both; in order for both genes to be turned on, an additional primer RNA is required. Implicit in the idea of stoichiometric control is the fact that the level of transcription of one gene is dependent upon the extent of transcription of other genes releasing

appropriate primers from their RNA precursors.

If one of the RNA primers shown in Figure 31 were to be repeatedly cleaved from the 5' ends of a whole series of mRNA precursor molecules whose synthesis it had primed one after another, this would be another way of switching genes on catalytically. Such a mechanism might be useful in cases where maintenance of a given level of a substance was desirable. However, one characteristic of the RNA processing reactions already discussed in detail is that the recognition sites extend beyond the cleavage sites. Recognition sites such as those shown in Figure 25 for cleavage of tyrosine tRNA include regions on both sides of the cleavage sites which end up separated from each other. If the 41-base extra fragment became linked to some other sequence, parts of the signal for processing would be missing and cleavage could not occur. Similarly, once a primer such as those shown in Figure 31 had been incorporated at the 5' terminus of a new mRNA precursor molecule, it is evident that an enzyme of the same specificity would no longer recognize any cleavage site which could release its 3' end; thus the primer sequence and its adjoining bases would be subject to the same RNA breakdown pattern which governs the majority of hnRNA sequences. The identical circumstance is observed in the case mentioned above, where mutants of E. coli $tRNA^{Tyr}$ so lower the rate of RNase P processing that the tRNA sequence produced in the cell, although stable after processing, is largely destroyed while part of the precursor, without ever passing through a tRNA stage. Thus as a rule we would view a single usage of a given primer as the most likely case.

The Extent of Primer-Initiated Transcription. Figure 32 illustrates one aspect of the stoichiometric control system discussed above. In order for coordination of gene expression of the sort shown in Figure 31 to occur, there must be at least two regions of the genome which encode the RNA primer--one within the gene which generates the precursor to the primer, and one in some other genetic unit, as shown. It is clear that an additional feature of this point of view is that a primer could cause another round of synthesis within the genetic unit

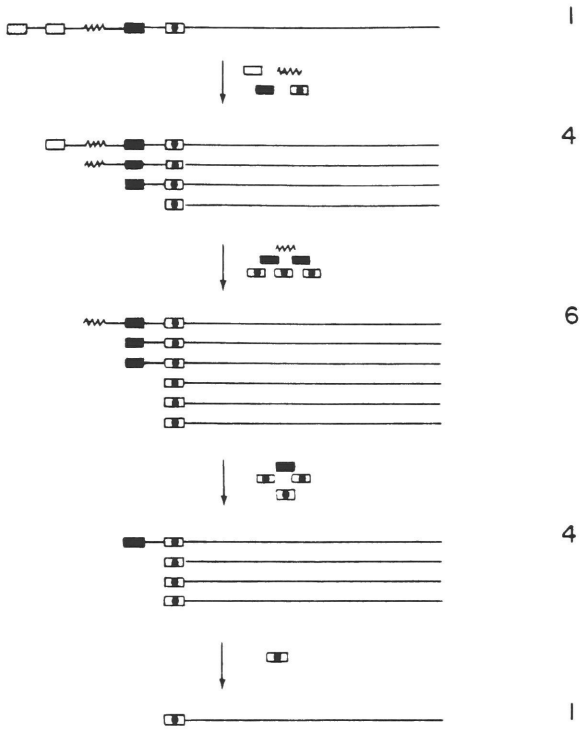


Figure 32

Transcription initiated by RNA primers as a way of controlling the extent of gene expression. All of the molecules shown schematically in this figure are mRNA precursors derived from the transcription of a single gene. The synthesis of the top molecule was primed by an RNA represented by (▨) which was presumably cleaved from some other mRNA precursor. Thus, this first transcript contains within its non-coding portion four other regions which yield RNA primers upon processing: (▭), (∿), (■) and (◧). As discussed in the text, it is likely that a primer at the 5' terminus could not be reused because of its failure to be processed correctly. In the special case depicted here, the four primers released from the first transcript each prime the transcription of another copy of the same gene, but with different start points within the non-coding region. As shown, these four new mRNA precursors of different lengths (but all containing the same mRNA region) contain a total of six internal primer sequences. We assume in this case that (i) primers at 5' termini are in general not reused, (ii) all internal primer sequences can be released to prime new synthesis, and (iii) at this stage of development there are no other sites within the genome available for competition for these particular primers. There will then be a total of 16 copies of this gene transcribed as a result of the introduction of the first primer (▨). Furthermore, the gene will be turned off at this point because of the absence of additional primers.

coding it. Thus in addition to inter-genic coordination, it is possible to obtain a burst of synthesis of the mRNA for a particular gene product which is controlled in both duration and extent by the location of the RNA primer which initiated synthesis of the first mRNA precursor. In Figure 32, then, we see a particular mRNA precursor which was initiated at a point so that four potential primer regions have been transcribed in the extra region between the 5' terminus and the coding region. In the extreme case shown here, where none of the primers turn on other genes, we see that at most 15 additional mRNA molecules will be synthesized before transcription of this particular genetic unit is automatically shut down. A key factor in determining this extent of transcription is our assumption that the processing events which generate primers are too subtle to allow re-use of a primer region once it has been located at the 5' end of an mRNA precursor.

Limitations. Two properties of such a system cannot be resolved simply by considering well characterized prototypes from other biological systems. The first is the problem of whether the functioning of RNA primers in this and other systems is concentration-dependent, or whether specific mechanisms exist which can efficiently utilize a single molecule to signal the desired event. Second, we cannot estimate the proportion of the extra RNA needed for such control systems. In one extreme, all 20,000 extra bases of an hnRNA molecule could consist of potential primers and recognition sites for processing enzymes. On the other hand, there may be other functions encoded in this extra regions which occupy a majority of its bases.

Finally, it seems that, in this kind of regulatory scheme, one event could follow another in rapid succession within the nucleus of the cell under consideration so that all of the events of differentiation could take place within a short time. In other words, the system lacks a pace-maker. A related problem is that no provision has been made for maintenance of stage of development through cell division. It may be that the specific population of small RNA's present within

the nucleus has one unique opportunity during each cell cycle to scrutinize the entire genome. In particular, during DNA replication, the RNA could bind to complementary regions within DNA, both at sites which had received primers for transcription during the life of that cell, and at other sites which had been inaccessible. Then the DNA would regain its double-stranded structure and fold up into condensed chromosomes for cell division with this population of RNA fragments built in at specific sites. When the nuclei of the daughter cells reform and transcription begins, these built-in RNA fragments could function as the first primers for transcription, and in this way establish the state of differentiation of the daughter cell. Studies already carried out on the close association of RNA with chromosomes could be re-evaluated in terms of this suggestion.

Economy. It is clear that by combining the specificity of RNA processing with the feature of base pairing found in primed nucleic acid synthesis reactions, one can obtain efficient utilization of the genetic information. Cleavage of a variety of substrates of unrelated sequence but related structure by a single processing enzyme could release a variety of unique fragments. Therefore it is not necessary to invoke a large number of RNA processing activities to yield the very large number of different RNA primer sequences required for control of this sort. That the number of such RNase activities in eukaryotes may be small is further suggested by findings of Altman et al. (1974) that two specific RNase activities (RNase P and RNase NU) appear to reside both in human tissue culture cells and E. coli.

Perhaps the most singular feature of this suggested scheme of gene control is the economy gained by using the genetic material itself, in the form of RNA primers, to carry out gene coordination. Schemes based entirely upon regulatory protein-nucleic acid interaction require not only genes from which the regulatory protein mRNA may be transcribed and the machinery by which it is translated, but also require mechanisms to control the expression of these genes, and so on.

Multiple Initiation Sites for RNA Transcription. As can be seen in Figure 31, depending on the primer used for the transcription of a particular mRNA precursor, transcripts would be expected to have sizes differing from each other, on the average, by multiples of the average distance between priming sites. When viewed in the perspective of a differentiating system, where transcription of a particular gene might have to be coordinated with the expression of different combinations of genes at different times, it seems likely that the priming sites within the gene could be ordered in a way that would accomplish this end. For example, in the relatively undifferentiated embryonic cell, most mRNA precursors might be of maximum length thus including all possible primers. As a cell becomes more specialized, transcription of the mRNA precursors for these specialized functions may be primed closer and closer to the coding region resulting in shorter mRNA precursors and fewer primers for coordinated expression of the lower number of more specialized genes. Thus, one might expect to find that the mRNA precursor for a given genetic unit in embryonic cells is longer than that of precursor molecules for the same mRNA in highly differentiated cells. Also, mRNA precursors of lower organisms may be shorter than those of advanced life forms which undergo extensive and complex differentiation. In any case, if our suggested priming of RNA transcription occurs, then the precursors to mRNA would be heterogeneous with respect to size, even before any processing events occurred.

Repetitive Nature of DNA Regions Encoding Primers. With the exception of the extreme case in which a particular primer can only function at the site of its own transcription, all primers would be encoded in the genome at least twice, and perhaps thousands of times. To carry out coordination of gene expression in the proposed way, identical DNA regions would be distributed to a series of functionally related genes, whatever their location. In addition, recognition sites for processing enzymes surrounding the primer regions might also be expected to be repetitious. Thus the class of DNA which carries the

primer sequences and the information for specific RNA processing might be expected to behave like the "middle-repetitive" DNA in kinetic studies of DNA renaturation.

The actual length of specified RNA sequence required for recognition by processing enzymes and for priming is unknown. It may be, for example, that there are secondary and tertiary structural requirements for processing enzyme recognition with very few restrictions at the primary sequence level and that the primer itself is relatively short. In this case, it is clear that the two RNA molecules which are processed by the same enzyme, yield the same primer and are primed in the same fashion, could be encoded in two DNA regions which would each behave as single copy DNA in most renaturation experiments.

Alternatively, it may be that RNA processing sites are often encoded by regions of DNA capable of intra-strand hydrogen bonding to form hairpin structures. Because HeLa cell hnRNA has been found to contain at least three percent of its nucleotides in such structures this is not an unreasonable consideration (Jelinek et al., 1973; Jelinek and Darnell, 1972; Ryskov et al., 1973). Regions of this type would behave like the most highly reiterated DNA in hybridization studies since such unimolecular reactions are not concentration dependent.

Possible Structures for Determining Priming Sites. The simplest way for an RNA primer to select its site of action would be for the entire length of DNA to be potentially available for interaction with a freely diffusing RNA fragment. Such passive priming would not require any special structures within the DNA. The minimum size for a uniquely specified RNA primer would increase with genome size. For example, E. coli would require a sequence 11 bases long in order to determine a single unique site of action for a molecule which recognizes base sequence while for a human genome 17 bases are required (Britten and Roberts, 1969).

Aside from the increasing length of primer required for specific recognition, passive priming may also depend upon maintenance of a

certain minimum concentration of a given primer in order for one RNA fragment to be effective. Thus instead of having one-to-one interactions between pathways, this idea would confine the effects of one gene's expression to an increase in the probability of expression of other genes, rather than making the expression of other genes a certainty.

The number of available priming sites within the genome could be limited by special structures within the DNA. For example, as suggested by Crick (1971) the DNA could be tightly folded in such a way that certain regions of double helix are opened up to allow recognition by regulatory elements. Such single-stranded regions would provide obvious sites for the formation of base-paired regions with RNA fragments. Alternatively, available priming sites might be limited by a mechanism whereby RNA fragments are carried by DNA-binding proteins which preferentially recognize certain regions of the DNA. These proteins could be RNA polymerase subunits. Conjectures of this sort about control of transcription in higher organisms will take on more meaning when more is known about the RNA polymerases which carry out various sorts of eukaryotic transcription. For example, although most nuclei are now thought to contain several polymerases which may transcribe different regions of the genome (Cold Spr. Harb. Symp., 1970), little is known about their in vivo template specificity or their detailed requirements for initiation. Preliminary experiments indicate that in several cases a polymerase thought to be associated with mRNA production in vivo has a marked preference for single-stranded DNA templates (Chambon et al., 1970; Miyagi and Wang, 1974; Roeder and Rutter, 1969; Stein and Hausen, 1970). Furthermore, although recent experiments have shown that such eukaryotic polymerases will incorporate gamma-^{32}P-labeled ATP or GTP into short RNA chains in vitro (Meihlac and Chambon, 1973; Miyagi and Wang, 1974) nothing is known about the extent of this initiation mechanism in vivo, nor about the ability of these enzymes to utilize primers.

If it is true that there are special structures within the DNA which somehow specify priming sites, this would introduce more flexibility into the possible modes by which gene coordination could be carried out. For example, much of the DNA could be tied up at any given time in structures (Crick, 1971) which prevent its potential hydrogen bonding with RNA fragments, rendering potential sites for primer action inaccessible.

Genetics of Primed Transcription Units. If gene coordination is mediated at the level of transcription such that RNA primers are synthesized in covalent linkage to precursors of mRNA from structural genes, then DNA elements encoding primer regions of mRNA precursor molecules are crucial to a number of pathways. Then, as Judd has suggested, "Mutations in these elements may act in pleiotropic fashion" (Judd and Young, 1973). Pleiotropic mutants have in fact been described in both Drosophila (Hadorn, 1965; Judd et al., 1972; Judd and Young, 1973; Kaufman et al., 1973; Poulson, 1945) and maize (McClintock, B., 1965, 1967). Furthermore, single point mutations which are lethal are found in unexpectedly high proportion of the Drosophila genes analyzed (e.g. 14 out of 16 in the zeste-white region of the X-chromosome (Judd et al., 1972). At the same time, Drosophila mutants which correlate with disappearance of the enzymatic activity in question are often not lethal. Mutations in regulatory elements, particularly if such elements are co-transcribed with the mRNA for the protein product of the gene in question, could explain these observations. A point mutation in a regulatory element as part of an mRNA precursor but situated outside the coding region could be lethal because the correct signals for gene coordination could then fail to be produced. In light of our knowledge of the profound effect of point mutations upon the rate of processing of RNA precursors in E. coli (as discussed earlier) it seems quite possible that a similar change in an RNA processing signal, causing either slow production or more rapid degradation of the RNA control element, could account for the effects discussed here.

Involvement of RNA Processing in Control at Levels Other than Transcription. The constant production of specific RNA fragments by processing enzymes could also influence other cellular events. For example, it is widely believed that RNA primers are active in the initiation of DNA replication (Sugino et al., 1972; Verma et al., 1971). Thus, replication at a particular site in the DNA could be influenced by any event which produces an RNA of the appropriate sequence. In particular, transcription of viral DNA could lead, during processing of the RNA transcript, to the production of RNA breakdown products which mimic DNA replication control elements. The possibility that such interference in host DNA replication by viruses such as polyoma and SV40 would merit special consideration.

Translation could also be influenced by the production of specific RNA fragments in the cytoplasm, which might occur during mRNA breakdown, for example. Such a fragment from one mRNA molecule could hydrogen bond with a region of a second mRNA molecule coding for a different protein, and create a double-stranded region. Cleavage of this region could cause the second mRNA to be broken down also; thus the appearance and turnover of a new mRNA molecule might trigger the destruction of another longer-lived mRNA which had previously been stable. Alternatively, interactions such as this could be used to activate previously latent mRNA's which would require cleavage to be converted into their active form.

RNA fragments might also be involved in protein chain initiation. Short RNA molecules can be co-purified with reticulocyte initiation factors, and their removal inactivates protein synthesis stimulated by these factors (Bogdanovsky et al., 1973; Heywood, 1970; Heywood et al., 1974; Schreier and Staehelin, 1965; Wigle and Smith, 1973). Production of different such RNA fragments at different times could thus be one way to vary the specificity of translation.

EXPERIMENTAL STATUS OF THIS HYPOTHESIS

Development of the point of view presented in this chapter leads to several lines of experimentation that might otherwise not have been established. It has had a direct effect upon our own research program as follows.[1]

1. The application of in vitro iodination to RNA species not previously available at high specific radioactivities has been studied (Chapter I; Robertson, Dickson, Model and Prensky, 1973).

2. Studies on the mechanism of action of the E. coli RNA processing enzyme, RNase III, are being carried out (unpublished experiments).

3. The complete nucleotide sequence of a small RNA species (F5) cleaved from the bacteriophage T7 early mRNA precursor by RNase III has been determined (Chapter III; Robertson, Dickson and Dunn, 1976, manuscript in preparation).

4. Sequence analysis of six greatly enriched oligonucleotides within the double-stranded regions interspersed in HeLa cell heterogeneous nuclear RNA has been carried out (Chapter III; Robertson, Jelinek and Dickson, 1976, manuscript in preparation).

5. The study of viroids has been initiated by applying the methods of in vitro iodination, RNA two-dimensional fingerprinting and sequencing analysis in order to characterize these biologically active molecules as unique and identifiable RNA species of low complexity (Chapter II; Dickson, Prensky and Robertson, 1975).

6. The possibility that viroid RNA is modified by its host is being investigated (Chapter II).

---

[1] The point of view presented in this chapter has also had an impact at a theoretical level. We were invited to apply these ideas to a consideration of potential roles for RNA in cellular development for presentation at the Fourth Conference on Embryonic and Fetal Antigens in Cancer. Our highly speculative response to this invitation is presented in Appendix III.

If viroid RNA represents some sort of escaped version of a regulatory molecule, studies designed to reveal the modes of viroid replication and pathogenicity could constitute valuable prototypes for the investigation of RNA molecules with potential regulatory roles in normal cells. Although the central issues of the hypothesis presented in this chapter--(i) that specific processing of RNA occurs in eukaryotic nucleic and (ii) that special RNA's cleaved from mRNA precursors contribute to the control of gene expression perhaps by priming the transcription of related genes and/or controlling the onset of DNA replication--may not be tested directly for some time, we expect these ideas will continue to stimulate experimentation in a variety of productive areas.

APPENDIX I

CALCULATION OF PERCENTAGE IODINATION FROM DPM PER MICROGRAM

Information required:

(i) amount of RNA in the iodination reaction

(ii) counts per minute ^{125}I incorporated

(iii) efficiency of ^{125}I counting

Assumptions involved in the calculation:

(i) all of the input RNA is recovered

Calculation of the percentage of C's iodinated from the specific activity:

1. The decay of ^{125}I follows first order kinetics resulting in the formation of one mole of Tellurium for every ^{125}I disintegration.

$$^{125}I \xrightarrow{k_1} Tu$$

2. The rate of disappearance of ^{125}I can then be expressed as

$$-\frac{dC}{dt} = kC$$

where "C" represents the concentration of ^{125}I.

3. Integrate:

$$-\ln C = kt + \text{constant of integration}$$

4. Evaluate at $t = 0$ and $C = C_o$ to determine the value of the constant of integration:

$$-\ln C_o = k(o) + \text{constant}$$

Therefore the constant $= -\ln C_o$

5. Substitute this value into equation (3):

$$-\ln C = kt - \ln C_o$$

or $$\ln C_o - \ln C = kt$$

or $$\frac{\ln C_o}{C} = kt$$

6. Find a value for k by letting $t = t_{1/2}$ and $C = C_o/2$:

$$\ln \frac{C_o}{C_o/2} = kt_{1/2}$$

$$\ln 2 = kt_{1/2}$$

$$k = \frac{\ln 2}{t_{1/2}}$$

7. Write equation (5) in the exponential form:

$$\frac{C_o}{C} = e^{kt}$$

i.e., $C = C_o e^{-kt}$

8. Differentiate with respect to time:

$$\frac{dC}{dt} = -kC_o e^{-kt}$$

9. Evaluate at $t = 0$:

$$\frac{dC}{dt} = -kC_o e^{o} \qquad \left[k = \frac{\ln 2}{t_{1/2}}\right]$$

$$\frac{dC}{dt} = -\left(\frac{\ln 2}{t_{1/2}}\right) C_o$$

$$C_o = \frac{dC}{dt}\left(\frac{t_{1/2}}{\ln 2}\right)$$

$$\boxed{C_o = \text{dpm}\left(\frac{t_{1/2}}{\ln 2}\right)}$$

Sample Calculation

(i) amount of RNA = 1 microgram

(ii) 10^8 cpm ^{125}I incorporated

(iii) efficiency of counting = 60% = 0.6

(iv) half-life of ^{125}I ($t_{1/2}$) = 60 days = 86,400 minutes

(v) RNA is 25% cytidylate residues

1. The number of ^{125}I atoms per microgram of RNA $= \frac{10^8}{0.6} \cdot \frac{86,400}{\ln 2}$

$= 2.08 \times 10^{13}$

2. The number of C's per microgram of RNA $= \frac{10^{-6}}{350} \cdot 6.023 \times 10^{23} \times 0.25$

$= 4.3 \times 10^{14}$ C's per microgram

3. Percentage of iodinated C's $= \frac{2.08 \times 10^{13}}{4.3 \times 10^{14}} \times 100$

$= 4.8\%$

APPENDIX II

THE BINOMIAL DISTRIBUTION

A. Use of the Binomial Distribution to Determine Whether Iodination is Random.

Let N = the no. of C's in an oligonucleotide = 7

n = the no. of iodo-C's in an oligonucleotide $\leq$ 7

f = the frequency of labeling = 28%

Then the proportion of fragments which have n of their N C's labeled is as follows:

$$\mathbb{P}(n) = \frac{N!}{n!(N-n)!} \cdot f^n(1-f)^{N-n}$$

n	$\frac{N!}{n!(N-n)!}$	f^n	$(1-f)^{N-n}$	$\mathbb{P}(n)$	% of labeled fragments	Relative contribution to cpm	% of cpm
1	7	.2800	.1393	.2730	30.344	30.34	13.92
2	21	.0784	.1935	.3186	35.414	70.82	32.49
3	35	.0220	.2687	.2069	22.997	69.00	31.65
4	35	.00615	.3732	.0803	8.925	35.72	16.38
5	21	.00172	.5184	.0187	2.078	10.40	4.77
6	7	.00048	.7200	.0024	0.267	1.62	0.74
7	1	.000135	1.0000	.000135	0.015	0.105	0.05
			Totals:	0.8997	100.000	218.005	100.00

The fraction of fragments with none of their C;s iodinated may be calculated as follows:

$$\mathbb{P}(o) = 1 - \sum_{n-1}^{7} \mathbb{P}(n) = 1 - 0.8997 = 0.1003$$

Thus, when 28% of the C's are converted to iodo-C, 10.03% of the fragments containing 7 C's have no iodo-C and hence no ^{125}I label.

B. Use of the Binomial Distribution to Relate CPM to Percentage Iodination

(i) Let %(n) be the percentage of the total cpm recovered in the spot corresponding to the oligonucleotide with n of its N C's iodinated.

Let a_n and a_{n+1} be the cpm recovered in two adjacent spots.

$$\frac{\%(n)}{\%(n+1)} = \frac{a_n}{\Sigma a_n} \times 100 \div \frac{a_{n+1}}{\Sigma a_n} \times 100$$

$$= \frac{a_n}{a_{n+1}}$$

(ii) From the Binomial Distribution expression

$$\frac{\%(n)}{\%(n+1)} = \frac{\frac{\mathbb{P}(n)}{\Sigma\mathbb{P}(n)} \times 100 \times n}{\Sigma \text{ Relative cpm}} \times 100 \div \frac{\frac{\mathbb{P}(n+1)}{\Sigma\mathbb{P}(n)} \times 100 \times (n+1)}{\Sigma \text{ Relative cpm}} \times 100$$

$$= \frac{\mathbb{P}(n)}{\mathbb{P}(n+1)} \cdot \frac{n}{(n+1)}$$

$$= \frac{N!}{n!(N-n)!} f^n(1-f)^{N-n} \div \frac{N!}{(n+1)!(N-(n+1))!} f^{n+1}(1-f)^{N-(n+1)}$$

$$\times \frac{n}{(n+1)}$$

$$= \frac{n}{(N-n)} \cdot \frac{(1-f)}{f}$$

(iii) Therefore $$\frac{a_n}{a_{n+1}} = \frac{n}{(N-n)} \cdot \frac{(1-f)}{f}$$

Thus, when a_n, a_{n+1}, N and n are known, f (frequency of iodination) can be calculated.

C. Sample Calculation of Frequency of Iodination from Binomial Distribution.

Consider the oligonucleotide CCAUACCACCCUG:

N = the total no. of C's = 7

n = no. of iodo-C's in a particular fragment

a_n = cpm in the spot with n of N C's iodinated = a_2 = 34,406 cpm

$a_{n+1} = a_3 = 33,798$ cpm

$$\frac{(1-f)}{f} = \frac{a_n}{a_{n+1}} \cdot \frac{(N-n)}{n} \quad \ldots . \text{ (see page 114)}$$

$$\therefore \frac{(1-f)}{f} = \frac{34,406}{33,798} \cdot \frac{(7-2)}{2} = 2.545$$

Plot $\frac{(1-f)}{f}$ as a function of f

$\frac{(1-f)}{f}$ goes from 99 to 1.0 as f goes from 0.01 to 0.5

Locate the value of f which gives $\frac{(1-f)}{f} = 2.545$

$f = 0.281$

Therefore, the percentage of C's iodinated, calculated from the Binomial Distribution, is 28.1%.

APPENDIX III

POTENTIAL REGULATORY ROLES FOR RNA IN CELLULAR DEVELOPMENT

In this discussion we have suggested a requirement for a line of communication between the cell surface and the genome so that regulatory events could occur in response to external signals. We have explored, in hypothetical terms, the possibility that this role could be carried out by RNA, and applied this reasoning to cells undergoing differentiation, including the immune cells. We have also suggested an important role for the RNA processing enzymes within the nucleus in selecting which subset of the RNA transcribed will become stable, and thus capable of action. Finally, these two ideas have been combined to show how such a system could account for normal or regenerative differentiation.

In Chapter III we have discussed the existence of populations of extra RNA in eukaryotic cell nuclei and specific RNA processing enzymes for their controlled metabolism. Figure 33 is a schematic representation of five possible roles for such RNA fragments. In this Appendix we will deal mainly with the suggestion that RNA cleaved from a transcript of one gene could prime the transcription at a second site in the genome thereby causing coordinate production of these gene products. In particular, we would like to suggest how phase-specific RNA molecules could be involved in maintenance of the differentiated state and how imbalance in the environment of the cell could lead to a scrambling of the program expressed.

A variety of proteins synthesized in embryonic cells have been observed to reappear in tumor cells (Alexander, 1972; Manes, 1974). Also present in these cells must be the corresponding phase-specific mRNA's. Within the context of the ideas about RNA precursors and specific processing of these molecules outlined above, we suggest that there are also phase-specific populations of RNA cleavage products confined to the nucleus which could function as regulatory entities.

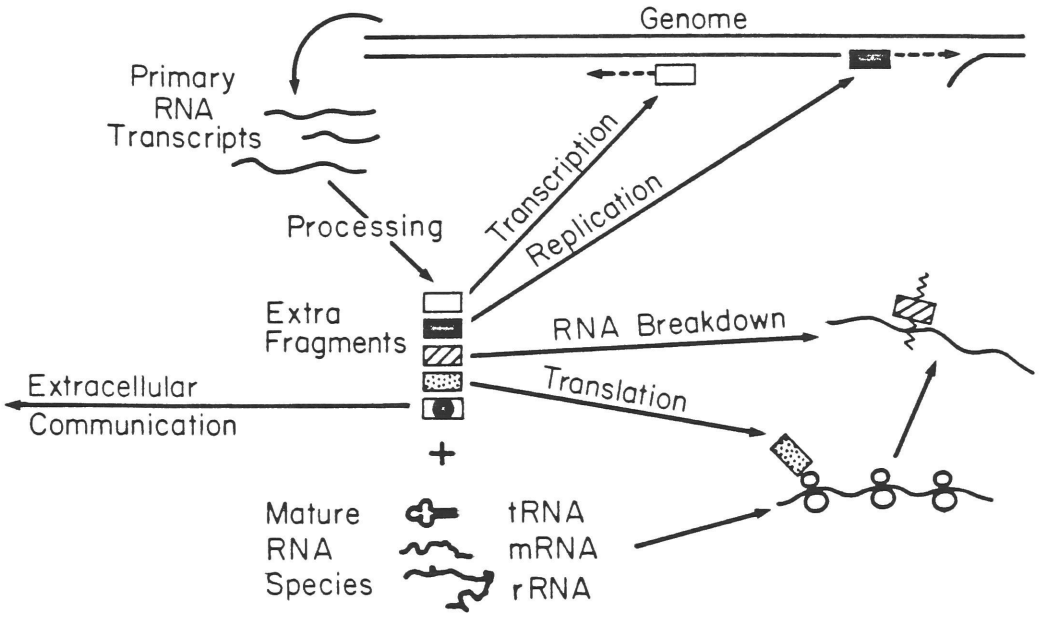


Figure 33

RNA processing as a source of regulatory RNA's. Here the primary RNA transcripts, consisting of the mature RNA's covalently linked to precursor-specific "extra" nucleotides, are processed by specific enzymes to yield the mature RNA species (tRNA, mRNA and rRNA) and also a group of specific RNA fragments from within the "extra" regions of the precursor RNA molecules. These specific fragments, represented by the symbols ▭ , ▬ , ▨ , ▩ and ◧ , are shown participating in five different cellular events: transcription, DNA replication, translation, initiation of RNA breakdown, and extracellular communication.

One way in which such regulatory RNA could be involved in the control of transcription during normal development is depicted in Figure 31. Small RNA fragments cleaved from the extra regions of the precursor to a particular mRNA would locate complementary regions within the genome, become engaged in hydrogen bonding, and act as primers in those locations for transcription of the adjacent region which would contain a precursor to a new mRNA (see legend to Figure 31 for details). As discussed in Chapter III this control mechanism would be stoichiometric (the RNA primer would be used up in the reaction which it initiates; see Figure 32) and could account for a burst of synthesis at a particular site which would be controlled both in extent and duration by an initial priming event. Details of these arguments will not be reviewed here. It is sufficient to say that within the context of control of transcription by proteins revealed in a series of bacterial studies (Jacob and Monod, 1961), the use of RNA as an additional control element would add flexibility, efficiency and elegance to a logical system of gene control.

State of the Genome in Cellular Development

In applying these ideas to questions of cellular development, we have made the assumption that no irreversible changes occur in the DNA of any cell undergoing differentiation. We realize that available data about immune cells are most often interpreted as favoring irreversible changes in the DNA by somatic mutation or recombination (Burnet, 1959; Edelman and Gally, 1967; Smithies, 1967) of developing lymphocytes resulting in production of one and only one antibody per cell (Burnet, 1959). However, any system which is to be a valid prototype for differentiation must allow cells to retain their full genetic complement. Such retention of full genetic capability has been demonstrated in nuclei of frog intestinal epithelium cells (Gurdon, 1968) and by the frequent observation that tumor cells from one terminally differentiated tissue type suddenly express a number of genetic characteristics normally associated with an unrelated cell type (Alexander, 1972).

A conservative geneticist's view would require that the genome remain unaltered throughout differentiation. It will soon be possible to determine experimentally whether or not irreversible changes do occur in the DNA of lymphocytes by combining the techniques of RNA:DNA hybridization with comparative fingerprinting analysis of antibody mRNA's. An exploratory application of such techniques to a particular mouse myeloma light chain mRNA has recently appeared (Rabbitts et al., 1975). By adding RNA fingerprinting analysis to the more conventional RNA:DNA hybridization approach, it was shown that RNA which hybridized to the single-copy DNA contained fragments from both the Variable and Constant regions of the mRNA. In addition, comparative RNA fingerprinting studies, which are capable of revealing the extent of variation among closely related RNA molecules (Robertson and Jeppesen, 1972), could be effectively employed in combination with the above approach (Rabbitts et al., 1975).

The somatic mutation hypothesis invokes random changes in the DNA to explain the enormous number of different antibodies that can be observed. This hypothesis predicts differences in the nucleic acid sequences of mRNA's coding for antibodies of identical amino acid sequence isolated from different animals. Myeloma lines have been independently isolated which secrete antibodies with light chains of identical sequence (Dayhoff, 1972). Isolation and sequence analysis of significant portions of light chain mRNA's from such cell lines should therefore confirm or refute the validity of the somatic mutation hypothesis.

For the balance of this discussion we shall make the assumption that the information content of the DNA remains intact, and thus a cell retains the potential of re-expressing embryonic genes even after becoming highly differentiated.

Signals for cellular development. If there are no irreversible changes in DNA during differentiation, then all cells are identical with respect to their DNA content. Thus some other component must be

directing and maintaining the state of differentiation. We would expect a particular state of differentiation to be maintained actively by regulatory machinery. All of the DNA would then be accessible to reactivation by changes in this pattern of regulation. Based on studies of cell-cell interactions it seems that there may be a line of communication between the macromolecules on cell surfaces and the DNA (Garber and Moscona, 1972; Greaves and Janossy, 1972; Moscona, 1975). Apparently cells, depending on their environment, can be triggered to undergo various sorts of growth patterns (Lehtonen et al., 1975; Toivonen et al., 1976). For example, consider the growth characteristics of adult and regenerating liver cells, where a change in cell environment drastically changes growth patterns from static to exponentially dividing (Bucher and Malt, 1971; Leduc, 1964). Perhaps the clearest example of a differentiating system requiring signals at the cell surface is the immune system. The information contained by a particular antigen is transmitted to a lymphocyte in such a way as to elicit the production of a particular protein composed of two different polypeptides, the heavy and the light chain, which together form a structure that binds with high affinity to the antigen in question.

Consider the possibility that RNA acts as an information carrier in this proposed line of communication between the cell surface and the DNA. This proposal can then be combined with our earlier suggestions about ways in which RNA could act as a regulatory element within the nucleus (Chapter III; Figure 31). In order to retain a particular signal for later action RNA synthesized early in development could remain complexed to a cell surface receptor protein awaiting a subsequent induction event. Such a hypothetical scheme might, for example, consist of (i) interaction of the inducing agent with the cell surface receptor protein, (ii) release of the latent RNA molecule from association with the receptor protein, and (iii) interaction of this released RNA with the DNA to bring about a change in gene expression. This series of events is depicted schematically in Figure 34. The mode of

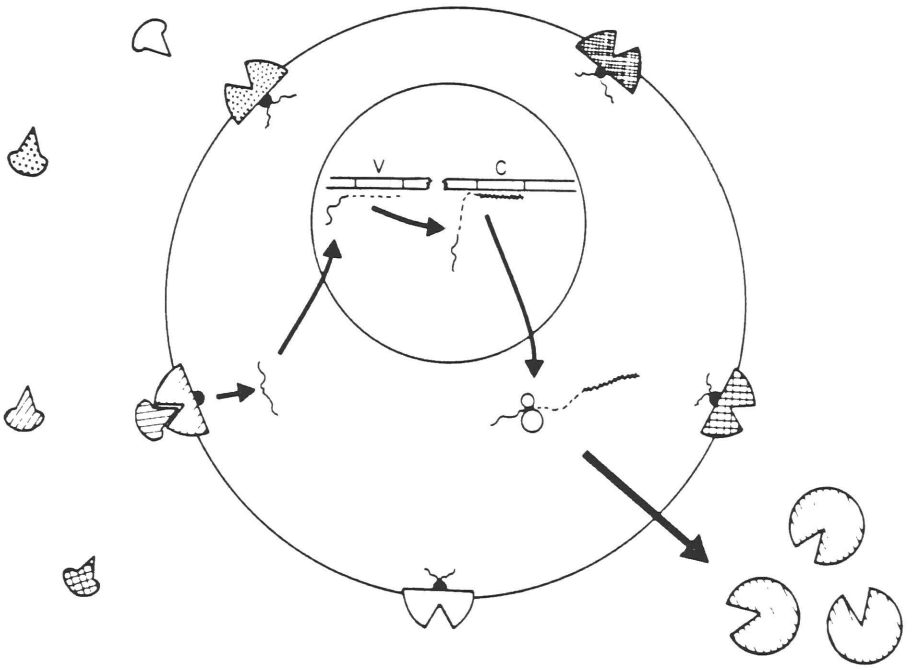


Figure 34

Hypothetical cell surface signals for differentiation. This is a schematic representation of a cell with a variety of receptors on its surface (⩙). The cell contains a nucleus in which is depicted two regions of double-stranded DNA containing the genes marked "V" and "C". Outside the cell are a variety of substances (△) which are capable of specific interaction with matching cell surface receptors (i.e., the dotted ones can only interact with the dotted receptors). A series of hypothetical events would allow this cell to respond to an encounter with a particular substance (▨) by a change in state of differentiation which would allow the production of the new proteins shown in the lower right hand corner of the figure as circles with wedges cut out of one side (◍). This change would be accomplished by the following series of events: (i) the inducing agent (▨) interacts with the cell surface receptor (⩙), (ii) the surface receptor undergoes a change which causes it to release an RNA fragment (~) which had been bound to the surface receptor since the time when this protein was synthesized. (iii) the RNA fragment travels to the nucleus where it locates a complementary region within the genome and acts there to prime the transcription of the adjacent region of the genome--namely the "V" gene, (iv) this RNA transcript locates a second region within the genome which is complementary to its 3' end and again acts as a primer for the continued transcription so that gene "C" is transcribed as a continuation of gene "V", (v) this RNA transcript travels to the cytoplasm where it is translated to produce the new protein (◍), (vi) in this example, the final step shown is the secretion of this new protein.

action of the regulatory RNA represented in this figure is that of priming of RNA transcription to initiate the production of RNA from a previously silent gene (see Figure 31).

If we apply this kind of thinking to the response of the immune system to a particular antigen, it leads to several further proposals. First, lymphocytes could contain an array of light chain variable regions (V regions) complexed to some of the "extra" RNA from the V region mRNA on their surfaces. When the cell is confronted by a particular antigen, if V regions with the ability to bind with high affinity to this antigen existed on a particular cell, these V region receptors would bind antigen, and at the same time, would release their fragment of latent RNA. This RNA would travel to the nucleus, locate the region from which it was originally transcribed, engage in complementary base pairing, and act as a primer for the transcription of the adjacent region--namely, the V region which can bind with high affinity to the antigen which triggered the reaction. Extending this logic to the next step, this RNA transcript of the DNA encoding the V region could contain, at its 3' end, a short region complementary to the DNA sequence immediately adjacent to the DNA encoding the light chain constant region (C region). Thus, a second priming event could occur resulting in the production of a complete light chain mRNA, accomplishing the joining of these regions without the requirement of scrambling the DNA in an irreversible manner (Edelman and Gally, 1967; Smithies, 1967).

Such a selection mechanism would allow lymphocytes to be multipotent with respect to their potential to respond to a variety of antigens thus obviating the requirement for a separate line of cells, each with only one specificity, for responding to every antigen.

In addition to considering informed signals involving RNA at cell surfaces mediating communication with the nucleus, we would like to suggest a second way in which major changes in gene expression could arise in differentiated cells. This second mechanism involves the

highly specific RNA cleaving enzymes capable of producing RNA signals from primary transcripts in the nucleus. If we make the assumption that members can be added to or deleted from this group of enzymes at specific stages of development, it would lead to the scheme of events shown in Figure 35. At early times, both genetic units A and B are expressed. Three specific RNA processing enzymes are present: Ez 1, Ez 2 and Ez 3 (as shown in Figure 35). At late times, the descendants of this cell no longer transcribe embryonic gene A, have lost Ez 1 and Ez 3, and have acquired a new RNA processing enzyme, Ez 5. In the absence of Ez 1 and Ez 3, two primers which were previously stabilized by maturation cleavage are now turned over. At the same time, the presence of a new enzyme (Ez 5) allows the stabilization of a regulatory RNA from the extra region of gene B that had previously been unstable. Thus, the altered state of differentiation is reflected in an altered population of RNA fragments within the nucleus.

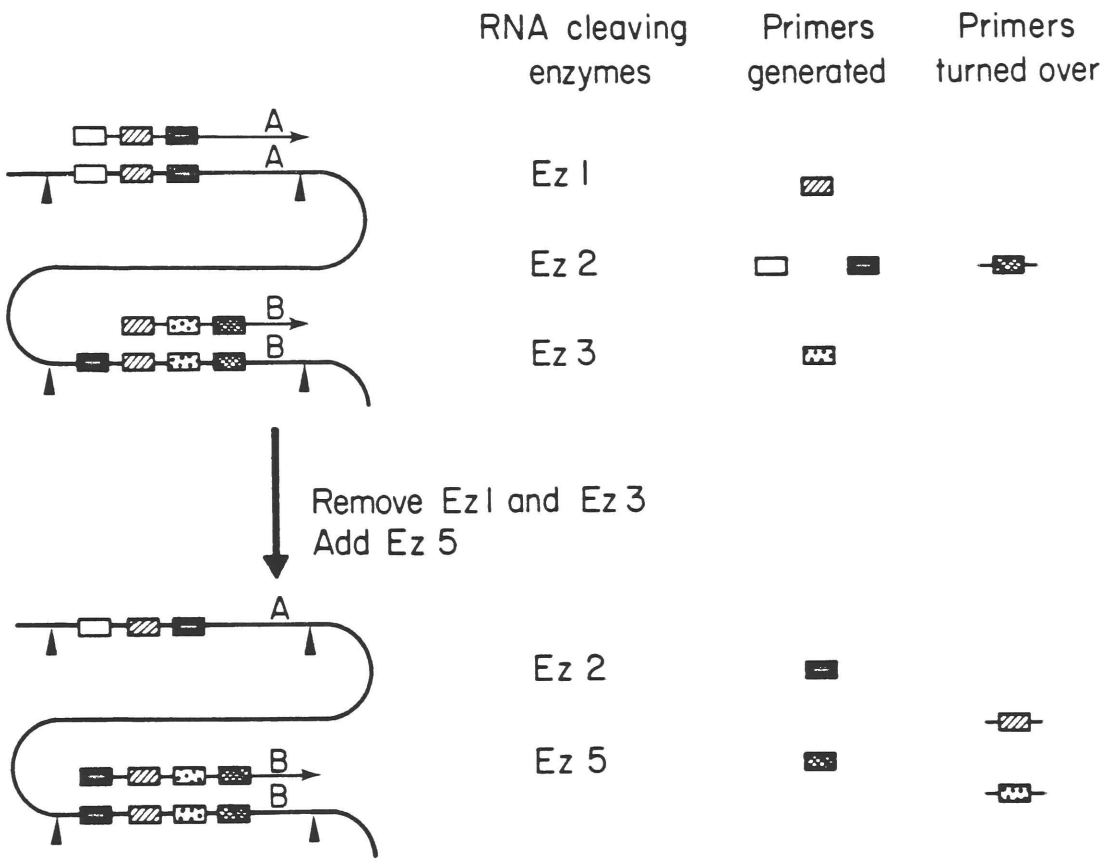
Now it is possible to discuss what would happen if Ez 1 were to reappear in the more highly differentiated cell. A regulatory RNA within the extra region of gene B that has not been released as a stable species since early in development is once more produced. This same RNA is capable of binding to the complementary region within the embryonic gene A and causing its re-expression. Furthermore, transcription of gene A would in turn lead to the production of a set of RNA regulatory elements which could prime coordinate transcription of further genes abnormal for this state of differentiation. Thus a scrambling of the genetic program of differentiation would have been accomplished.

But what is the proposed mechanism of reappearance of enzyme Ez 1? As mentioned previously, it is possible to think of highly differentiated cells as being actively maintained in a particular state by constant positive control at the level of transcription. Cells within organs such as liver retain the capacity to detect the absence of an intact organ and react by changing their state of differentiation to that of a developing liver cell and then repeat the developmental

Figure 35

Role of enzymes in maturation of hypothetical regulatory RNA fragments. The rectangular symbols (e.g., ▨) within the continuous folded lines at the left side of the figure represent those regions within the DNA that are destined to be transcribed into RNA and processed by specific enzymes to yield RNA primers for the control of transcription at some time during the development of some cell line. The solid folded line represents the DNA, in which the arrows mark the boundaries of two particular transcription units designated as "A" and "B". The template activity of the genome shown in the upper left region of the figure represents the transcription of both gene A and gene B during some early embryonic stage. The RNA processing enzymes present at this particular stage (Ez 1, Ez 2 and Ez 3) can cleave these large RNA transcripts in such a way as to release mature forms of four out of five potential primer molecules depicted (▨, ▭ , ▬ and ▩). The other such region (▩) which is transcribed but not processed in the appropriate way (because of the absence of the appropriate enzyme) undergoes degradation by less specific "scavenging" enzymes. We suggest that the four RNA fragments which undergo proper maturation are engaged in the control of transcription at this stage (see Figure 31). The template activity of the genome of an adult stage cell of this same cell line is depicted in the lower left portion of the figure. During development, synthesis of enzymes Ez 1 and Ez 3 has ceased and a new RNA processing enzyme, Ez 5, has appeared. Gene A represents an embryonic gene which is no longer transcribed. Two potential regulatory regions from within the transcript of gene B which were stabilized during the embryonic stage (▨, ▩) have now become unstable because of the absence of RNA processing enzymes Ez 1 and Ez 3. On the other hand, the presence of Ez 5 allows the stabilization of a fragment of the gene B transcript which was turned over during the embryonic stage (▩). Thus, the particular group of RNA processing enzymes present in the nucleus at each stage of development could have a central role in selecting the subpopulation of "extra" RNA which would undergo maturation within the nucleus to yield regulatory RNA molecules to take part in the control of transcription.

A key feature of the suggested role of RNA processing enzymes in the generation of regulatory RNA molecules is that, simply by the re-introduction of an embryonic processing enzyme into an adult stage nucleus, it would be possible to stabilize some RNA regions which prime the transcription of embryonic genes. These genes, in turn, would contain regulatory RNA's within their extra regions which would go on to prime the transcription of still other genes. Thus, an irreversible scrambling of the genetic program of that cell line would have been accomplished.


RNA cleaving enzymes
Primers generated
Primers turned over
A
A
Ez 1
Ez 2
B
B
Ez 3
Remove Ez1 and Ez 3
Add Ez 5
A
Ez 2
B
B
Ez 5

pathway to regenerate a new liver. It therefore seems possible that adult liver cells have receptors on their surfaces which constantly monitor the immediate neighbors, and when part of the liver is missing, cells receiving the signal could undergo a series of events something like those depicted in Figure 34. Thus a latent RNA would be released by the cell surface receptor, and travel to the nucleus. In keeping with our suggestion that major changes of state are mediated by changes in RNA processing enzymes, this RNA could prime the transcription of a gene encoding a new RNA processing enzyme. The presence of this enzyme could have profound effects upon the make-up of the population of stable RNA species within the nucleus (see Figure 35), and in turn, upon the state of differentiation of the cell.

Abnormal Differentiation

The authors were requested to consider whether the conceptual framework presented here could also shed light on the cause of abnormal differentiation such as that observed in the tumor state. The following speculations are therefore intended to stimulate further consideration of these questions. Tumors can apparently arise spontaneously, can be induced by chemical agents, radiation or tumor viruses, or can be genetically programmed. Is there a common target for all of these agents? Suppose that chemical carcinogens disrupt the sort of surface receptors we have been discussing, causing release of the latent regulatory RNA. As we have outlined, this would lead to scrambling of the genetic program. In the case of the tumor viruses, a change in the host group of specific nuclear RNA processing enzymes (as discussed in Chapter III in the case of Adenovirus) can now be seen as potentially capable of similarly profound effects on host gene expression. Radiation could act at the level of DNA disruption causing uncoupling of previously coordinated events. Finally, tumors with high heritability would seem to be the result of an error in the genetic instructions for a developmental program.

While highly speculative, the advantage to the various ideas outlined here is that, whether or not our suggested role for RNA in cellular development eventually gains experimental support, such ideas should serve to stimulate concrete experiments which will contribute to a solid basis for understanding whatever systems are eventually found to guide the expression of eukaryotic genes.

REFERENCES

Alexander, P. (1972). Nature 235, 137-140.

Altenburg, E. (1946). Am. Natur. 80, 559-567.

Altman, S. (1975). Cell 4, 21-29.

Altman, S., Bothwell, A. L. M. and Stark, B. C. (1974). Brookhaven Symp. in Biol. 26, 12-25.

Altman, S. and Robertson, H. D. (1973). Mol. Cell. Biochem. 1, 83-93.

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

Attardi, G., Parnas, H., Hwang, M. I. H. and Attardi, B. (1966). J. Mol. Biol. 20, 145-182.

Bagnall, R. H., Singh, R. P. and Clark, M. C. (1973). Canada Agriculture, Issue of Spring 1973.

Barrell, B. G. (1971). In Procedures in Nucleic Acid Research, G. L. Cantoni and D. R. Davies, eds., Vol. 2, pp. 751-779, Harper and Row, New York.

Beckwith, J. and Zipser, D., eds. (1970). The Lactose Operon, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Bekhor, I., Kung, G. M. and Bonner, J. (1969). J. Mol. Biol. 39, 351-364.

Bogdanovsky, D., Hermann, W. and Schapira, G. (1973). Biochem. Biophys. Res. Commun. 54, 25-32.

Bothwell, A. L. M. and Altman, S. (1975). J. Biol. Chem. 250, 1451-1459.

Brandhorst, B. P. and McConkey, E. H. (1974). J. Mol. Biol. 85, 451-463.

Britten, R. J. and Davidson, E. H. (1969). Science 165, 349-357.

Britten, R. J. and Roberts, R. B. (1969). Carnegie Inst. Wash. Yearb. 69, 113-124.

Brownlee, G. G. and Sanger, F. (1969). Eur. J. Biochem. 11, 395-399.

Brownlee, G. G., Sanger, F. and Barrell, B. G. (1968). J. Mol. Biol. 34, 379-412.

Bucher, N. L. R. and Malt, R. A. (1971). Regeneration of Liver and Kidney, Little, Brown and Co., Boston.

Burnet, F. M. (1959). In The Clonal Selection Theory of Acquired Immunity, Vanderbilt and Cambridge University Presses, Nashville.

Chambon, P., Gissinger, F., Mandel, J. L., Jr., Kedinger, C., Gniazdowski, M. and Meihlac, M. (1970). Cold Spring Harbor Symp. Quant. Biol. 35, 693-707.

Cold Spring Harbor Symp. Quant. Biol. (1970). 35.

Commerford, S. L. (1971). Biochemistry 10, 1993-2000.

Crain, W. (1975). Fourth Conference on Embryonic and Fetal Antigens in Cancer, in press.

Crick, F. H. C. (1971). Nature 234, 25-27.

Dahlberg, J. E., Harada, F. and Sawyer, R. C. (1975). Cold Spring Harbor Symp. Quant. Biol. 39, 925-932.

Darnell, J. E., Jr. (1968). Bacteriol. Rev. 32, 262-290.

Davidson, E. H. and Britten, R. J. (1973). Quart. Rev. Biol. 48, 565-613.

Davies, J. W., Kaesberg, P. and Diener, T. O. (1974). Virology 61, 281-286.

Dayhoff, M. O., ed. (1972). Atlas of Protein Sequence and Structure, Vol. 5, The National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D. C.

Dickson, E. and Robertson, H. D. (1976). Cancer Research, in press.

Dickson, E., Prensky, W. and Robertson, H. D. (1975). Virology 68, 309-316.

Diener, T. O. (1971a). Virology 43, 75-89.

Diener, T. O. (1971b). Virology 45, 411-428.

Diener, T. O. (1972). Virology 50, 606-609.

Diener, T. O. (1973). Phytopathology 63, 1328-1329.

Diener, T. O. and Lawson, R. M. (1973). Virology 51, 94-101.

Diener, T. O. and Raymer, W. B. (1967). Science 158, 378-381.

Diener, T. O. and Raymer, W. B. (1969). Virology 37, 351-366.

Diener, T. O. and Raymer, W. B. (1971). In Descriptions of Plant Viruses, No. 66, Commonwealth Mycological Institute and Association of Applied Biologists.

Diener, T. O., Schneider, I. R. and Smith, D. R. (1974). Virology 57, 577-581.

Diener, T. O. and Smith, D. R. (1973). Virology 53, 359-365.

Diener, T. O., Smith, D. R. and O'Brien, M. J. (1972). Virology 48, 844-846.

Dorland's Illustrated Medical Dictionary (1965). W. B. Saunders Co., Philadelphia, Pa.

Duesberg, P. H. and Vogt, P. K. (1973). J. Virol. 12, 594-599.

Dunn, J. J. (1976). J. Biol. Chem. 251, 3807-3814.

Dunn, J. J. and Studier, F. W. (1973a). Proc. Nat. Acad. Sci. US 70, 1559-1563.

Dunn, J. J. and Studier, F. W. (1973b). Proc. Nat. Acad. Sci. US 70, 3296-3300.

Edelman, G. M. and Gally, J. A. (1967). Proc. Nat. Acad. Sci. US 57, 353-358.

Forget, B. G. and Weissman, S. M. (1967). Science 158, 1695-1699.

Franklin, R. M. (1966). Proc. Nat. Acad. Sci. US 55, 1504-1511.

Garber, B. B. and Moscona, A. A. (1972). Developmental Biology 27, 217-234.

Georgiev, G. P. (1969). J. Theoret. Biol. 25, 473-490.

Gierer, A. (1973). Cold Spring Harbor Symp. Quant. Biol. 38, 951-961.

Gilbert, W. and Müller-Hill, B. (1970). In The Lactose Operon, J. R. Beckwith and D. Zipser, eds., pp. 93-109, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Gilham, P. T. (1962). J. Amer. Chem. Soc. 84, 687-688.

Goss, R. W. (1926). Phytopathology 16, 299-303.

Greaves, M. F. and Janossy, G. (1972). In Cell Interactions, F. Silvestri, ed., North Holland Publishing Co.

Hadorn, E. (1964). Brookhaven Symp. Biol. 18, 148-161.

Hall, T. C., Wepprich, R. K., Davies, J. W., Weathers, L. G. and Semancik, J. S. (1974). Virology 61, 486-492.

Hausen, P. and Stein, H. (1970). Eur. J. Biochem. 14, 278-283.

Heywood, S. M., Kennedy, D. S. and Bester, A. J. (1974). Proc. Nat. Acad. Sci. US 71, 2428-2431.

Hollings, M. and Stone, O. M. (1973). Ann. Appl. Biol. 74, 333-348.

Holmes, D. S., Mayfield, J. E. and Bonner, J. (1974). Biochemistry 13, 849-855.

Hörst, R. K. (1975). Phytopathology 65, 1000-1003.

Huang, R. C. C. and Huang, P. C. (1969). J. Mol. Biol. 39, 365-378.

Ikemura, T. and Dahlberg, J. E. (1973). J. Biol Chem. 248, 5024-5032.

Jacob, F. and Monod, J. (1961). Cold Spring Harbor Symp. Quant. Biol. 26, 193-211.

Jacob, F. and Monod, J. (1961). J. Mol. Biol. 3, 318-356.

Jacobson, A., Housman, D. and Prensky, W. (1973). Molec. Biol. Reports 1, 209-213.

Jelinek, W., Adesnik, M., Salditt, M., Sheiness, D., Wall, R., Molloy, G., Philipson, L. and Darnell, J. E. (1973). J. Mol. Biol. 75, 515-532.

Jelinek, W. and Darnell, J. E. (1972). Proc. Nat. Acad. Sci. US 69, 2537-2541.

Jeppesen, P. G. N. (1971). Doctoral Thesis, University of Cambridge.

Judd, B. H., Shen, M. W. and Kaufman, T. C. (1972). Genetics 71, 139-156.

Judd, B. H. and Young, M. W. (1973). Cold Spring Harbor Symp. Quant. Biol. 38, 573-579.

Kaufman, T. C., Tasaka, S. E. and Suzuki, D. T. (1973). Genetics 75, 299-321.

Kleppe, R. and Khorana, H. G. (1972). J. Biol Chem. 247, 6149-6156.

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

Leduc, E. H. (1964). In The Liver: Morphology, Biochemistry, Physiology, C. Rouiller, ed., Vol. II, pp. 63-83, Academic Press, New York.

Lehtonen, E., Wartiova, J., Nordling, S. and Saxen, L. (1975). J. Embryol. Exp. Morphol. 33, 187-203.

Ling, V. (1972). J. Mol. Biol. 64, 87-102.

Lodish, H. F. and Zinder, N. D. (1966). J. Mol. Biol. 19, 333-348.

Manes, C. (1974). Cancer Res. 34, 2044-2052.

McClintock, B. (1965). Brookhaven Symp. Biol. 18, 162-184.

McClintock, B. (1967). Developmental Biology (Suppl.) 1, 84.

Meihlac, M. and Chambon, P. (1973). Eur. J. Biochem. 35, 454-463.

Miyagi, M. and Wang, T. Y. (1974). Int. J. Biochem. 5, 113-122.

Model, P. and Zinder, N. D. (1974). J. Mol. Biol. 83, 231-251.

Monahan, J. J. and Hall, R. H. (Jan. 1974). CRC Crit. Rev. Biochem., pp. 67-112.

Moscona, A. A. (1975). Fourth Conference on Embryonic and Fetal Antigens in Cancer, in press.

Nichols, J. L. (1970). Nature 225, 147-151.

Nichols, J. L. and Robertson, H. D. (1971). Biochim. Biophys. Acta 228, 676-681.

Paul, J. (1972). Nature 238, 444-446.

Philipson, L., Pettersson, U., Lindberg, U. and Tibbets, C. (1974). Cold Spring Harbor Symp. Quant. Biol. 39, 447-456.

Pieczenik, G., Model, P. and Robertson, H. D. (1974). J. Mol. Biol. 90, 191-214.

Poulson, D. F. (1945). Amer. Natur. 79, 340.

Prensky, W. (1976). In Methods in Cell Biology, D. M. Prescott, ed., Vol 13, pp. 121-152, Academic Press, New York.

Prensky, W., Steffensen, D. M. and Hughes, W. L. (1973). Proc. Nat. Acad. Sci. US 70, 1860-1864.

Rabbitts, T. H., Jarvis, J. M. and Milstein, C. (1975). Cell 6, 5-12.

Randles, J. W. (1975). Phytopathology 65, 163-167.

Raymer, W. B. and Diener, T. O. (1969). Virology 37, 343-350.

Robertson, H. D., Altman, S. and Smith, J. D. (1972). J. Biol. Chem. 247, 5243-5251.

Robertson, H. D. and Dickson, E. (1974). Brookhaven Symp. in Biol. 26, 240-266.

Robertson, H. D., Dickson, E., Model, P. and Prensky, W. (1973). Proc. Nat. Acad. Sci. US 70, 3260-3264.

Robertson, H. D. and Dunn, J. J. (1975). J. Biol. Chem. 250, 3050-3056.

Robertson, H. D. and Hunter, T. (1975). J. Biol. Chem. 250, 418-425.

Robertson, H. D. and Jeppesen, P. G. N. (1972). J. Mol. Biol. 68, 417-428.

Robertson, H. D., Webster, R. E. and Zinder, N. D. (1967) Virology 32, 718-719.

Robertson, H. D., Webster, R. E. and Zinder, N. D. (1968). J. Biol. Chem. 243, 82-91.

Roeder, R. G. and Rutter, W. J. (1969). Nature 224, 234-237.

Romaine, C. P. and Horst, R. K. (1975). Virology 64, 86-95.

Rosenberg, M., Kramer, R. A. and Steitz, J. A. (1974). Brookhaven Symp. in Biol. 26, 277-285.

Ryskov, A. P., Saunders, G. F., Farashyan, V. R. and Georgiev, G. P. (1973). Biochim. Biophys. Acta 312, 152-164.

Sanger, H. L. (1972). Advances in Biosciences 8, 103-116.

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

Scherrer, K and Marcaud, L. (1968). J. Cell. Physiol. 72, 181-212.

Schreier, M. H. and Staehelin, T. (1973). Nature New Biol. 242, 35-38.

Schultz, E. S. and Folsom, D. (1923) J. Agr. Res. 25, 43.

Semancik, J. S., Magnuson, D. S. and Weathers, L. G. (1973a). Virology 52, 292-294.

Semancik, J. S., Morris, T. J. and Weathers, L. G. (1973b). Virology 53, 448-456.

Semancik, J. S. and Weathers, L. G. (1972a). Nature New Biol. 237, 242-244.

Semancik, J. S. and Weathers, L. G. (1972b). Virology 47, 456-466.

Semancik, J. S. and Weathers, L. G. (1972c). Virology 49, 622-625.

Semancik, J. S., Morris, T. J., Weathers, L. G., Rodorf, B. F. and Kearns, D. R. (1975). Virology 63, 160-167.

Sharp, P. A., Gallimore, P. H. and Flint, S. H. (1974). Cold Spring Harbor Symp. Quant. Biol. 39, 457-473.

Singh, R. P. (1970). Amer. Potato J. 47, 225-227.

Singh, R. P. (1971). Phytopathology 61, 1034-1035.

Singh, R. P. (1973). Amer. Potato J. 50, 111-123.

Singh, R. P. and Bagnall, R. H. (1968). Phytopathology 58, 696-699.

Singh, R. P. and Clark, M. C. (1973). FAO Plant Prot. Bull. 21, 121-125.

Singh, R. P. and Finnie, R. E. (1973). Can. Plant Dis. Surv., Issued Sept. 1973.

Singh, R. P., Finnie, R. E. and Bagnall, R. M. (1970). Amer. Potato J. 47, 289-293.

Singh, R. P., Finnie, R. E. and Bagnall, R. M. (1971). Amer. Potato J. 48, 262-267.

Singh, R. P., Michniewicz, J. J. and Narang, S. A. (1974). Can. J. Biochem. 52, 809-812.

Singh, R. P. and O'Brien, M. J. (1970). Amer. Potato J. 47, 367-371.

Smithies, O. (1967). Science 157, 267-273.

Sogo, J. M., Koller, Th. and Diener, T. O. (1973). Virology 55, 70-80.

Stedman's Medical Dictionary (1961). Williams & Wilkins Co., Baltimore, Maryland.

Stein, H. and Hausen, P. (1970). Cold Spring Harbor Symp. Quant. Biol. 35, 709-717.

Sugino, A., Hirose, S. and Okazaki, R. (1972). Proc. Nat. Acad. Sci. US 69, 1863-1867.

Sutton, W. D. (1972). Nature New Biol. 237, 70-71.

Terao, T., Dahlberg, J. E. and Khorana, H. G. (1972). J. Biol. Chem. 247, 6147-6166.

Toivonen, S., Tarin, D. and Saxen, L. (1976). Differentiation 5, 49-55.

Tsanev, R. and Sendov, B. (1971). J. Theor. Biol. 30, 337-393.

Van Dorst, H. J. M. and Peters, D. (1974). Neth. J. Pl. Path. 80, 85-96.

Varricchio, F. and Prensky, W. Manuscript in preparation.

Verma, I. M., Meuth, N. L., Bromfield, E., Manly, K. F. and Baltimore, D. (1971). Nature New Biol. 233, 131-134.

Warner, J. R., Soeiro, R., Birnboim, H. C., Girard, M. and Darnell, J. E. (1966). J. Mol. Biol. 19, 349-361.

Watson, J. D. (1972). Nature New Biol. 239, 197-201.

Webster, R. E., Engelhardt, D. L. and Zinder, N. D. (1967). J. Mol. Biol. 29, 27-43.

Wigle, D. T. and Smith, A. E. (1973). Nature New Biol. 242, 136-140.

Williams, J. G. and Penman, S. (1975). Cell 6, 197-206.

Zubay, G., Schwartz, D. and Beckwith, J. (1970). Cold Spring Harbor Symp. Quant. Biol. 35, 433-435.

End