LBL-792

PARTIAL SEQUENCE ANALYSIS OF TOBACCO MOSAIC VIRUS RNA

Lily Wong Sun
(Ph. D. Thesis)

April 1972

AEC Contract No. W-7405-eng-48

For Reference

Not to be taken from this room

LBL-792

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

0 0 0 0 3 8 0 0 2 7 9

PARTIAL SEQUENCE ANALYSIS OF TOBACCO MOSAIC VIRUS RNA

Abstract

Lily Wong Sun

^{32}P-labeled TMV-RNA with specific activity of 10^5 CPM/μg has been obtained. Its complete T_1-RNase digest has been fractionated by a two-dimensional high voltage ionophoresis technique (Sanger *et al.*, 1965). From the resultant T_1-RNase fingerprint, the total chainlength of TMV-RNA has been determined to be 6,284 ± 2.4%. When carefully treated with sodium dodecyl sulfate, 1.19, 3.14, and 4.58% of the coat protein subunits have been removed from the 3' end of TMV. T_1-RNase fingerprints are obtained for the exposed end of TMV-RNA. By comparison of the recovery of the oligonucleotides from T_1-RNase digests at these three different extents of strippings, the Ω-mer (Lloyd & Mandeles, 1970) terminates 158 to 163 nucleotides from the 3' end of TMV-RNA. The T_1-RNase digestion products that occur prior to the Ω-mer are: 11-12 G, 3-4 CG, 5 AG, 2-3 UG, 2 CAG, 1 ACG, 1 AAG, 1 UCG, 2 CUG, 1 AUG, 1 UUG, 2 $(A_2,C)G$, 2 $(U,C_2)G$, 1 CAAAG, 1 $(A_2,C_2,U)G$, 1 $(A,C_2,U_2)G$, 2 $(A_2,C,U_2)G$, 1 $(A_3,U_2)G$, 1 $(C,U_3)G$, 1 $(A_2,C,U_3)G$, 1 $(A_3,C_4,U_2)G$, 1 $(A_4,C_3,U_5)G$, 1 $(A_2,C_2,U_5)G$, and 1 CCCA (Steinschneider & Fraenkel-Conrat, 1966).

For all products from T_1-RNase digests of the exposed 3' end of TMV-RNA, base composition and their molar yields have been obtained. Their nucleotide sequences have been studied by pancreatic-RNase digestion.

ACKNOWLEDGEMENTS

I wish to express my gratitude to Dr. Ignacio Tinoco, Jr., under whose guidance the work reported in this thesis was conducted. His advice and encouragement have made my years as a graduate student both fruitful and interesting. I shall always be indebted to Dr. Stanley Mandeles for his continuous interest and advice. The invaluable discussions with him greatly extended my knowledge of the RNA sequence analysis.

I have benefited from the association with the members of Dr. Tinoco's research group, and in particular Dr. Olke Uhlenbeck, Dr. Arlene Blum, Mr. Mark Levine, Mr. Philip Borer, Mr. Eric Wickstrom, Mrs. Carol Cech, Dr. Dana Carroll, Dr. Carl Formoso, Mrs. Barbara Dengler and Mrs. Sharon McConnell. I would also like to extend my thanks to Mr. Dave Garfin in Dr. Mandeles's group.

I would like to thank Dr. Melvin Calvin, in whose laboratory ^{32}P-TMV was prepared, for the generous gifts of inorganic ^{32}P. Mr. Edward Heilstad has provided constant radiation protection. Mr. Ronald Felzer has helped preparing some ^{32}P-TMV used in this work.

To my husband, Alex, and parents goes my deep appreciation for their many hours of patient listening and their unwavering encouragement.

0 0 0 0 3 8 0 0 2 8 0

Table of Contents

	Page
LIST OF ABBREVIATIONS	vi
I. INTRODUCTION	1
II. MATERIALS AND METHODS	6
A. Materials	6
B. Preparation of ^{32}P-labeled TMV	7
C. Extraction of ^{32}P-TMV-RNA	10
D. Hydrolysis of ^{32}P-TMV-RNA with Enzyme	13
(i) T_1-RNase	15
(ii) PAN-RNase	15
E. Partial Stripping and Hydrolysis of the Exposed Nucleic Acid	15
F. Two-dimensional High Voltage Ionophoresis Fractionation Procedure	17
(i) Apparatus used for ionophoresis	17
(ii) Buffer systems and markers	19
(iii) Ionophoresis materials	19
(iv) Spotting and the first dimensional ionophoresis	20
(v) Transferring and the second dimensional ionophoresis	22
(vi) Radioautography	24
G. Structure of the Oligonucleotides	24
(i) Elution	24
(ii) Base composition	27

Page

(iii) Sequence analysis 29

III. RESULTS 35

A. T_1-RNase Fingerprint of Whole ^{32}P-TMV-RNA 35

B. The Chainlength Determination of ^{32}P-TMV-RNA . . 38

C. SDS Polar Stripping of Protein Subunits from ^{32}P-TMV 42

(i) Stripping condition 42

(ii) The extent of stripping 46

(iii) The distribution of stripping 49

D. T_1-RNase Fingerprints of Exposed 3' Terminus of ^{32}P-TMV-RNA 55

(i) Structure of the oligonucleotides 60

(ii) Molar yields of the oligonucleotides 71

E. Location of the Ω-mer 75

IV. DISCUSSION 86

REFERENCES 94

0 0 0 0 3 8 0 0 2 3 1

LIST OF ABBREVIATIONS

RNA	Ribonucleic acid
TMV	Tobacco mosaic virus
TMV-RNA	Tobacco mosaic virus ribonucleic acid
PSV	Partially stripped TMV
PV	Enzyme digested PSV
T_1-RNase	Ribonuclease T_1
PAN-RNase	Pancreatic ribonuclease
EDTA	Ethylene-diaminetetraacetic acid, buffer adjusted to the given pH with sat. NaOH
DEAE-	Diethylaminoethyl-
SDS	Sodium dodecyl sulfate
TEA	Triethylamine
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol, buffer adjusted to the given pH with conc. HCl
A_{260nm}	An O.D. unit equivalent to the amount of material which will produce an absorbance of 1.0 at 260 nm when dissolved in 1 ml and measured in a 1 cm path length cell
A or A residue	Adenosine 3'-phosphate, a purine nucleotide
G or G residue	Quanosine 3'-phosphate, a purine nucleotide
C or C residue	Cytidine 3'-phosphate, a pyrimidine nucleotide
U or U residue	Uridine 3'-phosphate, a pyrimidine nucleotide
!	Cyclic phosphate

0 0 0 0 3 8 0 0 2 8 2

I. INTRODUCTION

Tobacco mosaic virus (TMV) is composed of a ribonucleic acid (TMV-RNA) and an outer coat of protein. TMV-RNA is a single stranded ribonucleic acid of 6,340 nucleotides. It winds in a loose helix with protein molecules packed between turns. Each turn has a diameter of 18.0 nm and a pitch of 2.3 nm. There are 49 protein molecules in every three turns. Each virus particle consists of 2,130 identical protein molecules each of which consists of 158 amino acids of known sequence. The helical arrangement of the ribonucleic acid and all the coat protein molecules forms a rod shaped virus of 300 nm in length and 18.0 nm in diameter (Fraenkel-Conrat, Molecular Basis of Virology).

Crystalline TMV was first isolated by Stanley in 1935 (Stanley, 1935). Its ribonucleoprotein nature was demonstrated in the following years. However, it was not until 1956 that Gierer & Schramm (1956), and independently Fraenkel-Conrat (1956), demonstrated that TMV-RNA alone, in the complete absence of protein, was necessary and sufficient to produce infection. These experiments definitely showed that the viral TMV-RNA carries the genetic specificity to code both for its own replication and for the amino acid sequence in its specific coat protein. Since then, much more attention was paid to the structural analysis of the ribonucleic acid.

The structural analysis of TMV-RNA would provide information as how this ribonucleic acid acts as a genetic materials, i.e. how this RNA is copied during replication and how the information is transferred into protein. A knowledge of the nucleotide sequence

of TMV-RNA would allow one to determine the structure of those sites on the RNA responsible for the control of replication and protein synthesis. An A residue was first reported at both the 3' and the 5' end of the TMV-RNA by Sugiyama & Fraenkel-Conrat (1963). The 3' terminal nucleotide sequence -GCCCA was later determined by Steinschneider & Fraenkel-Conrat (1966) using chemical stepwise degradation method. The same sequence was independently reported by Mandeles (1967) who used end group labeling technique as well as a series of chromatographic separations. To determine the complete nucleotide sequence of a ribonucleic acid of the size of TMV-RNA requires, however, an incredible amount of work. But, there are two unique properties of TMV which might facilitate partial sequence analysis.

The polar packing arrangement of the RNA and the protein molecules is one of the unique properties of TMV. From their reconstitution experiments, Stussi et al. (1969) have shown that the protein molecules build onto the RNA in a polar manner, i.e., starting from one end of the RNA; and Ohno et al. (1971) were able to conclude that this polar reconstitution begins at the 5' end of the RNA. This evidence is consistent with the polar stripping experiments performed by May & Knight (1965). They found that the coat protein molecules can be removed by sodium dodecyl sulfate (SDS) sequentially from the 3' end of the viral rod. The exposed 3' end of the RNA then becomes accessible to ribonucleases digestion, whereas in the intact virus, the RNA is protected from any action of ribonucleases by its coat protein. This should enable one to

determine the sequence of the exposed part of the RNA, for one is no longer dealing with a RNA of more than six thousand nucleotides.

The other unique property of TMV was found independently by Mundry (1969) and Mandeles (1968) in their sequence studies. Mundry reported a unique long segment lacking G residues in TMV-RNA. Upon chainlength separation of complete T_1-ribonuclease (T_1-RNase) digests of whole TMV-RNA, Mandeles noticed the presence of three long oligonucleotides. Chromatographic evidence indicated that each occurs only once along the whole RNA chain. The chainlength of these three unique fragments were determined by Lloyd (Ph.D. thesis, 1969). The longest fragment, called the Ω-mer, has a chainlength of 70 nucleotides (Lloyd & Mandeles, 1970). By employing the SDS polar stripping property Mandeles (1968) has been able to locate the Ω-mer within the first 200 nucleotides of the 3' end. The work reported here uses the Ω-mer as a marker to aid the sequence analysis of the 3' terminus of TMV-RNA; the location of this Ω-mer has also been studied.

Progress in the determination of nucleotide sequences in ribonucleic acid has been limited by the techniques available for the fractionation of oligonucleotides produced by enzyme hydrolysis. The methods most used in recent years are the two-dimensional procedure which employs ionophoresis and chromatography on paper (Rushizky & Sober, 1962; Armstrong _et al._, 1964), and ion-exchange column chromatography on DEAE-cellulose or DEAE-sephadex (Rushizky _et al._, 1964). The paper chromatographic technique is used for speed and ease of operation, whereas ion-exchange chromatography

is used for efficient separation. The advantages of both techniques may be combined by the use of ion-exchange paper. Sanger and his coworkers (Sanger *et al.*, 1965) described this method, which uses high-voltage ionophoresis in both dimensions, for the fractionation of ribonuclease digests of ^{32}P-labeled RNA. Since then the sequences of a number of transfer RNA's of 77-78 nucleotides have been determined (Holley *et al.*, 1965; RajBhandary *et al.*, 1967; Dube *et al.*, 1968). The sequence for the 5S component of 120 nucleotides of ribosomal RNA from *E. coli* has also been reported (Brownlee *et al.*, 1968). More recently, with the additional tools of homochromatography on thin layers of DEAE-cellulose (Brownlee & Sanger, 1969) and polyacrylamide gel electrophoresis (Peacock & Dingman, 1967; Adams *et al.*, 1969), a tremendous number of nucleotide sequences have been reported. Nucleotide sequences present within 16S ribosomal RNA from *E. coli* were determined by Fellner *et al.* (1970) and Ehresmann *et al.* (1970). The first 175 nucleotide sequence from the 5' terminus of Qβ RNA synthesized *in vitro* has been reported by Billeter *et al.* (1969). Adams *et al.* (1969) and Nichols (1970) have reported sequences present in the coat protein cistron and its termination region of R17 bacteriophage RNA respectively. The sequence of the three ribosomal binding sites in the same phage RNA has been reported by Steitz (1969). Moreover, the sequences on both ends of bacteriophage MS2 RNA have been reported by Fiers *et al.* (1969). The work reported here has been devoted to find further sequence information for the RNA of TMV, the best known plant virus.

P-0 0 0 3 8 0 2 8 4

In this work, the two-dimensional high-voltage ionophoresis fractionation technique has been used as the method for fractionating T_1-RNase digest of ^{32}P-labeled TMV-RNA. High-voltage ionophoresis on cellulose acetate strips at pH 3.5 has been used in the first dimension. One advantage of using cellulose acetate strips instead of paper is that it results in considerably less streaking of large oligonucleotides. However, this limits one to very small amounts of material. Thus ^{32}P-labeled TMV-RNA has been used in all the work. Oligonucleotides are detected and estimated by radioautography and scintillation counting technique. In the second dimension high-voltage ionophoresis has been carried out on DEAE-cellulose papers at pH 1.9. Fractionation on DEAE-paper is due both to ion-exchange and electrophoretic effects. The ionophoresis essentially opposes the ion-exchange effect as the more acidic components move faster by ionophoresis but slower by ion-exchange. The radioautograph of the resultant two-dimensional fractionation is known as the fingerprint. With the application of the SDS polar stripping property of the virus, as little as 1-5% of the coat protein molecules have been removed from the 3' end. Fingerprints were obtained for T_1-RNase digests of the exposed 3' terminus. The sequence study performed in this work includes the determination of base composition, nucleotide sequences and relative molar ratio of the oligonucleotides present in all the fingerprints.

II. MATERIALS AND METHODS

A. Materials

Cellulose acetate electrophoresis strips were obtained from Millipore Corp., Bedford, Mass. 01730.

DEAE-cellulose papers were Whatman Chromedia DE81 which were made in England by W. & R. Balston Ltd., through Reeve Angels in Clifton, New Jersey.

T_1-RNase was from Toyko, Japan.

Whatman 1 and 3 MM chromatography papers were purchased from Van Waters & Rogers, Inc.

Dyes used as markers to follow the course of an ionophoresis contained equal volumes of 1% Xylene Cyanol F. F (blue), 2% Orange G (yellow) and 1% Acid Fuchsin (pink); all were from George T. Gurr Ltd., London.

Tobacco plants were kindly provided by the plant pathology green house of the University; only Turkish variety was used in this work.

The grinder used to homogenize infected Tobacco leaves is a omni-mixer homogenizer from Ivan Sorvall Inc., Norwalk, Conn. It consists of a basic omni-mixer and a stainless steel leak-proof grinding chamber.

Millipore filters with 0.45μ grid membrane were obtained from Nalge Co., Rochester, N.Y.

Centrifuge tubes with screw on caps for model L30 rotor and L40 rotor were obtained from the Nalge Co., N.Y.

Agarosegel A-150m 100-200 mesh was obtained from Bio-Rad Lab., Richmond, Calif.

X-ray films (no screen) and their developer and fixer were obtained from Picker's X-Ray of S.F., Calif.

High-voltage electrophoresis equipment and coolant were from Savant Instrument Inc., Hicksville, N.Y.

B. Preparation of ^{32}P-labeled TMV

The detailed procedure for preparation of ^{32}P-labeled TMV is described in the following. It can be considered as a modified Mandeles and Bruening (1968) method. The Turkish variety of the Tobacco plant about one month old with ten to twelve leaves was used in all preparations. One plant was used for each preparation. The leaves were infected with a solution containing 0.1 mg/ml of unlabeled TMV, 7% carborundum as abrasive and 1% K_2HPO_4 (pH 8.2). Yarwood & Fulton (1967) have discussed in detail a variety of mechanics for infection. The brush method was found to be the most convenient. The inoculum with abrasive and phosphate were applied to the upper leaf surface by brushing with a wet paint brush. Carborundum makes microwounds on the leaf surface. Although wounds are necessary for infection, they heal quickly. So it is advisable to apply inoculum and abrasive simultaneously. Phosphate makes the infection more effective and has given the most increase in infection.

Leaves were detached after one day of inoculation. A total of 20 mCi of inorganic ^{32}P was immediately added to the petioles. Since each leaf was placed in one petri dish, the amount of ^{32}P should be distributed evenly among all the leaves. 3-5 drops of

^{32}P for each leaf was easily taken up in 15-20 minutes. Then an equal amount of H_2O was added. After all the H_2O was taken up the 'hot' leaves were carefully wrapped in a plastic bag and allowed to stand at room temperature for a week. All the leaves were well preserved in this incubation period since the moisture was well kept within the plastic bag. Six days were enough for the biosynthesis of complete virus rods and incorporation of ^{32}P (Commoner et al., 1962). At the end of the incubation the leaves containing ^{32}P-labeled TMV were left in a freezer overnight in the same plastic bag and with adequate shielding. All procedures involving radioactive ^{32}P were done in a closed chamber made of half inch lucite which shielded more than 90% of the radiation. Gloves were worn at all times to avoid direct contact of any radioactive materials. A 9-12 inch tweezer was used to carry the 'hot' materials.

The frozen 'hot' leaves were quite brittle and they could be easily crushed into fine pieces by means of hammering. The broken pieces were poured into the pre-chilled stainless steel Sorvall grinding chamber and ground for as long as the leaves were still brittle (2-3 minutes). By doing so the leaves were broken into even finer particles. Then 200 ml of cold 0.04 M EDTA buffer containing 5% $NaHCO_3$ at pH 8 was added into the grinding chamber. The mixture was ground at the homogenizer's full speed for fifteen minutes or until it was homogeneous. During the course of the grinding the chamber was lowered into a methoxyethanol dry ice bath kept at 0° to -5° C to prevent any temperature increase of the mixture.

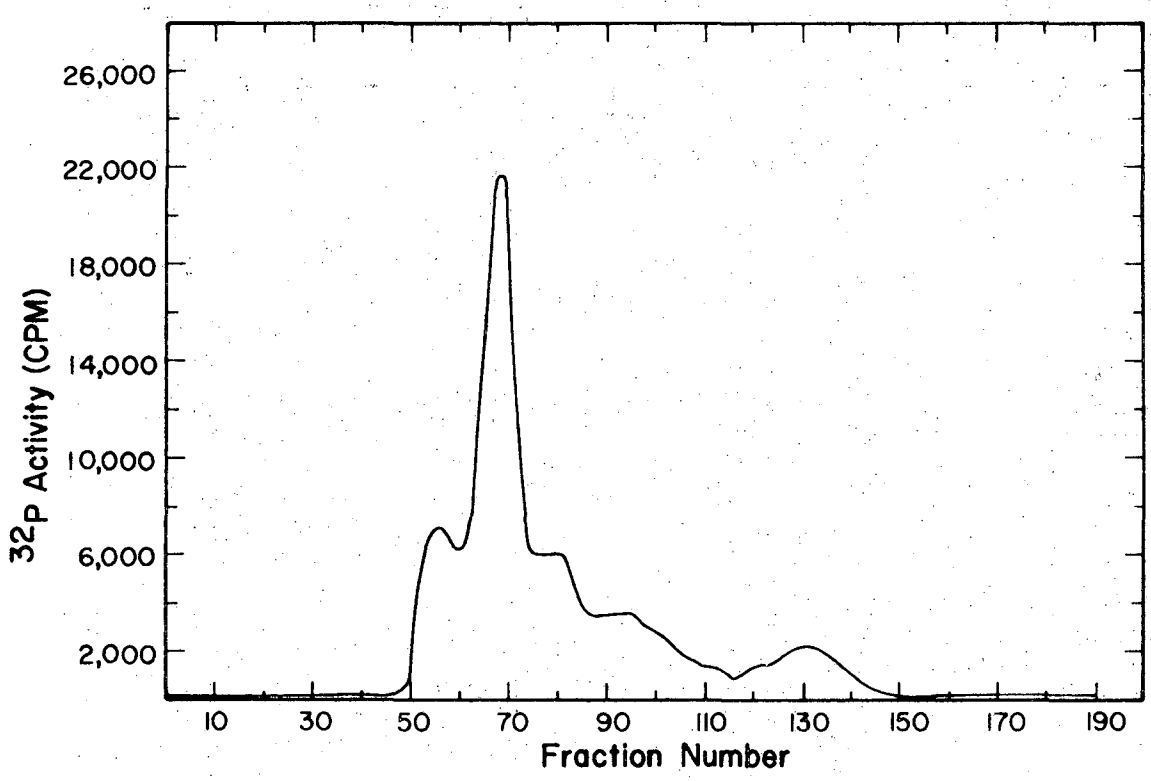
The thick, but aqueous, homogenate was centrifuged at 9,000 rpm and 0° C for 30 minutes in the GSA rotor of Sorvall centrifuge. The sedimented residue was discarded into radioactive waste can. The supernatant was saved, centrifuged again at 9,000 rpm and 0° C for 30 minutes to remove any remaining green residue. The supernatant from the second centrifuge was poured into model L30 centrifuge tubes. They were carefully balanced, capped, and centrifuged at 28,000 rpm and 0° C for 90 minutes. The dark brown supernatant solution was discarded into liquid radioactive waste can. The pellets containing ^{32}P-labeled TMV were covered with a total of 50 ml of 0.002 M EDTA-0.02 M cacodylate buffer (EDTA-cac buffer) pH 7.2 and allowed to stand in the cold room overnight. The centrifuge tubes were then swirled to disperse the virus evenly. The contents of the tubes were combined. After repeating the low speed and high speed centrifugations the virus pellet was suspended in 10 ml of cold EDTA-cac buffer pH 7.2 and again allowed to stand in the cold overnight. At this point the virus pellets were usually fairly translucent. The virus suspension was then filtered through a millipore filter with 0.45μ grid membrane. The filtrate was centrifuged at 28,000 rpm and 0° C for 90 minutes to bring the virus down to a pellet. The millipore filtration and high speed centrifugation were repeated, if necessary, until there was no discernible color in the virus pellet.

This pellet was then suspended in 1-2 ml of EDTA-cac buffer pH 7.2. The mixture was finally purified by passing through an agarose-gel A150 M column. The use of the millipore filter was to

remove any dimer and higher degree aggregated virus particles. The agarose-gel column separates viruses according to their sizes. Thus any short or aggregated virus rods were separated from the main portion of the 300 nm rods. The column was pre-equilibrated in EDTA-cac buffer pH 7.2. Elution was done with the same buffer at a flow rate of 4 ml per hour. A typical elution profile is shown in Fig. 1. Only those materials in the main peak centered around fraction 70 were taken, combined, passed through millipore filter, then centrifuged at 28,000 rpm again at 0° C for 90 minutes. The resultant pellet was suspended in the desired buffer for subsequent experiments. The yield of the ^{32}P-labeled TMV was estimated from its absorption at 265 nm. 1 mg of TMV is equivalent to 3.06 A_{265nm} units. Its activity was counted in a scintillation counter. The yields obtained from the above procedures are summarized in Table I.

C. Extraction of ^{32}P-TMV-RNA

The viral RNA was separated from its coat protein by the phenol extraction method in the presence of bentonite to prevent degradation by extraneous nucleases (Singer & Fraenkel-Conrat, 1961). 5-10 mg/ml of ^{32}P-TMV in 0.01 M Tris buffer pH 7.4 was stirred for 10 minutes with an equal volume of distilled phenol which was saturated with 0.01 M Tris buffer pH 7.4. 0.2 mg/ml of bentonite was used during the course of stirring. Centrifugation at 2,000 rpm for 10 minutes in Sorvall ss-34 rotor separated the mixture into two phases with bentonite as a layer at the interface between these phases. The upper aqueous phase was carefully transferred into a container



XBL 7110-1575

Fig. 1. ^{32}P-TMV Elution Profile

Resin: Agarose-gel A150 M 100-200 mesh.
Solvent: 0.002 M EDTA-0.02 M Cac buffer, pH 7.2.
Bed volume: 300 ml (2 X 100 cm column).
Sample load: 2 ml (10-20 mg of ^{32}P-TMV)
Flow rate: 4 ml/hr.

Table I

Yields of ^{32}P-TMV

Preparation	1	2	3	4
Amount of ^{32}P used in mCi	14	20	20	20
g of infected leaves	65	55	55	30
mg of ^{32}P-TMV	19.75	10.90	10.48	5.52
Activity of ^{32}P-TMV counts/min (CPM)	29×10^6	55.6×10^6	59.5×10^6	29.10×10^6
% incorporation	0.250	0.324	0.295	0.132
Specific activity CPM/mg of ^{32}P-TMV	1.47×10^6	5.10×10^6	5.68×10^6	5.27×10^6

which already contained about 1 mg of bentonite. Care must be taken not to remove any of the interface material with the aqueous phase. An equal volume of phenol was added. Stirring, centrifugation and separation of phases were carried out as before. The aqueous phase was centrifuged at 10,000 rpm in the ss-34 rotor for 30 minutes to remove any remaining bentonite. The supernatant was then extracted twice with 2X volume of ether to remove any contaminated phenol. After removal of ether by suction the aqueous solution containing ^{32}P-TMV-RNA was dialyzed against 95% ethanol for 2 hours. This was done instead of precipitating the RNA with ethanol. The volume of the dialysis tubing can then be decreased by making an extra knot. It was followed by dialysis against 0.001 M Tris buffer pH 7.4 to remove ethanol and excess salt. This method is modified from that described by Mandeles & Bruening (1968). The yield of RNA was estimated from its absorption at 260 nm. Its activity was counted in a scintillation counter. ^{32}P-TMV-RNA with specific activity as high as 1.0×10^5 CPM/μg has been obtained. The yields obtained from various preparations are summarized in Table II.

D. Hydrolysis of ^{32}P-TMV-RNA with Enzymes

In order to obtain reproducible results, the RNA must be completely hydrolyzed at all specific sites. Cyclic nucleotides must also be eliminated since they will lead to a duplication of each spot in a fingerprint. Too vigorous conditions should not be used since splitting of bonds other than the specific sites will occur. Therefore it is advisable to digest with relatively high

Table II

Yields of ^{32}P-TMV-RNA

Preparations	1	2	3	4
mg of ^{32}P-TMV	19.75	2.16	1.75	0.91
A_{260mu} of ^{32}P-TMV-RNA	24.40	3.25	1.58	1.07
Activity of ^{32}P-TMV-RNA counts/min (CPM)	22.90×10^6	9.44×10^6	6.38×10^6	3.9×10^6
Specific activity				
CPM/A_{260nm}	0.94×10^6	2.90×10^6	4.03×10^6	3.65×10^6
CPM/μg*	0.24×10^5	0.73×10^5	1.01×10^5	0.91×10^5
% yield	86.8	93.5	72.5	93.5

* 1 mg of RNA = 25 A_{260nm} units.

concentrations of enzyme for a short time. To avoid any contaminants of ribonucleases, gloves were required and all glassware used was washed in hot detergent and/or chromic acid, thoroughly rinsed, dried, then covered with aluminium foil and baked for 12 hours at 180-200° C.

(i) T_1-RNase:

Enzyme to substrate ratio used was 2 enzyme units (Egami & Takahashi, 1961) for every A_{260nm} unit of RNA. For relatively large amount of RNA, i.e. in the order of mgs, incubation was done in 1 ml solution at 40° C for 4 hours. With 200 μg or less, incubation was carried out in 10-20 μl in a sealed drawn-out melting point tube at 37° C for 4 hours. The digest in the capillary can be directly applied to the cellulose acetate strip for fractionation.

(ii) Pancreatic ribonuclease (PAN-RNase):

Hydrolysis by PAN-RNase, in all cases, was carried out in drawn-out melting point tubes due to the small amount of materials. For exact conditions used please refer to section G (iii).

E. Partial Stripping and Hydrolysis of the Exposed Nucleic Acid

The absorbance at 400 nm of a suspension containing 4-8 mg/ml of ^{32}P-TMV and 20-40 μg/ml of bentonite in EDTA-cac buffer pH 7.2 was measured. To this was added 25 μl/ml of 20% SDS so that the final concentration in SDS was 0.5%. This suspension was quickly

well mixed in the 1 cm path length cell and its absorbance at 400 nm was read immediately. The coat protein was being removed by SDS in a stepwise fashion from the 3' end of the viral rod (May & Knight, 1965). The extent of this polar stripping was followed by observing the scattering at 400 nm. The stripping was quenched with a large amount of cold buffer when the desired amount of protein had been removed. For example, to obtain a 5% stripping, the reaction mixture was diluted at least tenfold with cold EDTA-cac buffer when the absorbance at 400 nm had decreased 5%. This diluted suspension was then centrifuged in the L30 rotor at 28,000 rpm and 0° C for 3 hours. The supernatant was discarded. The pellet was resuspended in 10 ml of EDTA-cac buffer containing 20 μg/ml of bentonite. To ensure complete removal of SDS and the stripped protein, the suspension was recentrifuged at 28,000 rpm and 0° C for 3 hours. Then the pellet was covered with 1 ml of 0.001 M tris buffer pH 7.4 containing 20 μg/ml of bentonite and allowed to stand overnight in the cold.

The partially stripped virus (PSV) was freed of bentonite by centrifugation at 10,000 rpm for 30 minutes in the L30 rotor. The exposed 3' end of the RNA was hydrolysed with 2-3 enzyme units of T_1-RNase at 40° C for 4 hours. At the end of the digestion 100 μg of bentonite was added and the mixture was centrifuged in the L40 rotor at 37,000 rpm and 0° C for 3 hours. The partial virus (PV) formed a pellet at the bottom of the centrifuge tube. The supernatant, containing the nucleotides from T_1-RNase digestion, was recentrifuged at 37,000 rpm and 0° C for 3 hours to ensure complete removal of the PV. This supernatant was freeze-dried. The products

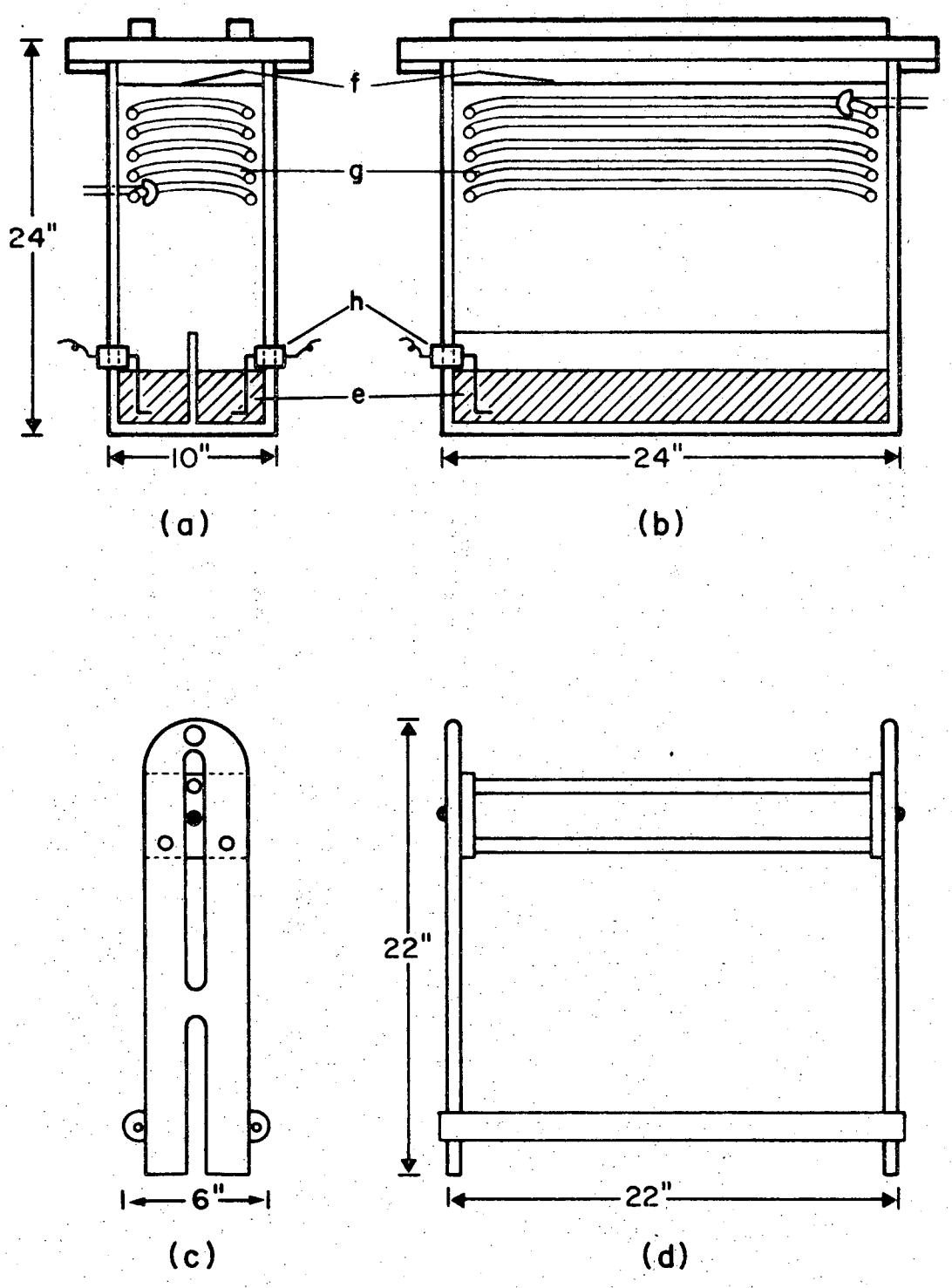
were separated by the two-dimensional high-voltage ionophoresis as described in the following section. The PV pellet obtained from the two centrifugations was resuspended in a minimum amount of EDTA-cac buffer. Its ^{32}P activity together with that of PSV and the T_1-RNase digestion products gave the precise extent of stripping.

F. Two-Dimensional High-Voltage Ionophoresis Fractionation Procedure

This fractionation procedure was originally described by Sanger, Brownlee & Barrell (1965). It was used throughout this work for separating all T_1-RNase digestion products of ^{32}P-labeled TMV-RNA. Therefore the detail experimental procedures used are worth describing in the following.

(i) Apparatus used for ionophoresis:

Ionophoresis tank and rack suitable for cellulose acetate strips and DEAE-papers are described in Fig. 2. The tank and rack are made of half inch lucite, all joints are glued. The anode and cathode compartments are each filled with 6-7 liters of buffer as shown in (e). The whole tank is then filled with Varsol to the level (f). Varsol, an electrophoresis coolant is used to keep the paper from drying out while placed in the tank. Cold water is passed through the stainless steel cooling coil (g) during each run. Electrodes (h) are of platinum coated wires. They are connected to a power supply capable of producing up to 3,000 volts with a current range of 0-300 mamp.



XBL 7110-1573

Fig. 2. Ionophoresis tank and rack

(a) Front of tank, (b) Side of tank,
(c) Front of rack, (d) Side of rack.

0 0 0 0 3 8 0 0 2 9 0

(ii) Buffer systems and markers:

pH 3.5 buffer consists of 0.5% pyridine and 5% acetic acid (v/v). This was used with Whatman No. 1 papers for base composition studies and with DEAE-paper for secondary enzyme digestions.

pH 3.5 buffer in 7 M urea was prepared by titrating a solution of 5% acetic acid in freshly prepared 7 M urea with redistilled pyridine. This buffer was used with cellulose acetate strips in the first dimensional ionophoresis. pH of this buffer changes upon standing. Glacial acetic acid should be added whenever necessary to bring the pH back down to 3.5. The increase of pH is due to the decomposition of urea into ammonia. Therefore it is advisable to check the pH of this buffer and make necessary adjustments before each run.

pH 1.9 buffer consists of 7% formic acid. Ionophoresis on DEAE-paper in this solution was used in the second dimension.

The dye mixture which is used as ionophoresis markers consists of equal volumes of 1% xylene cyanol F.F (blue), 2% Orange G (yellow) and 1% Acid fuchsin (red).

(iii) Ionophoresis materials:

Cellulose acetate strips were used in the first dimension. This material is very brittle when dry but easy to handle when wet. Fractionation in the second dimension was carried out on sheets of DEAE-paper. This

material on the other hand is easy to handle when dry, but very fragile when wet. Care must be taken when handling fragile materials.

The following two sections should be referred to Fig. 3.

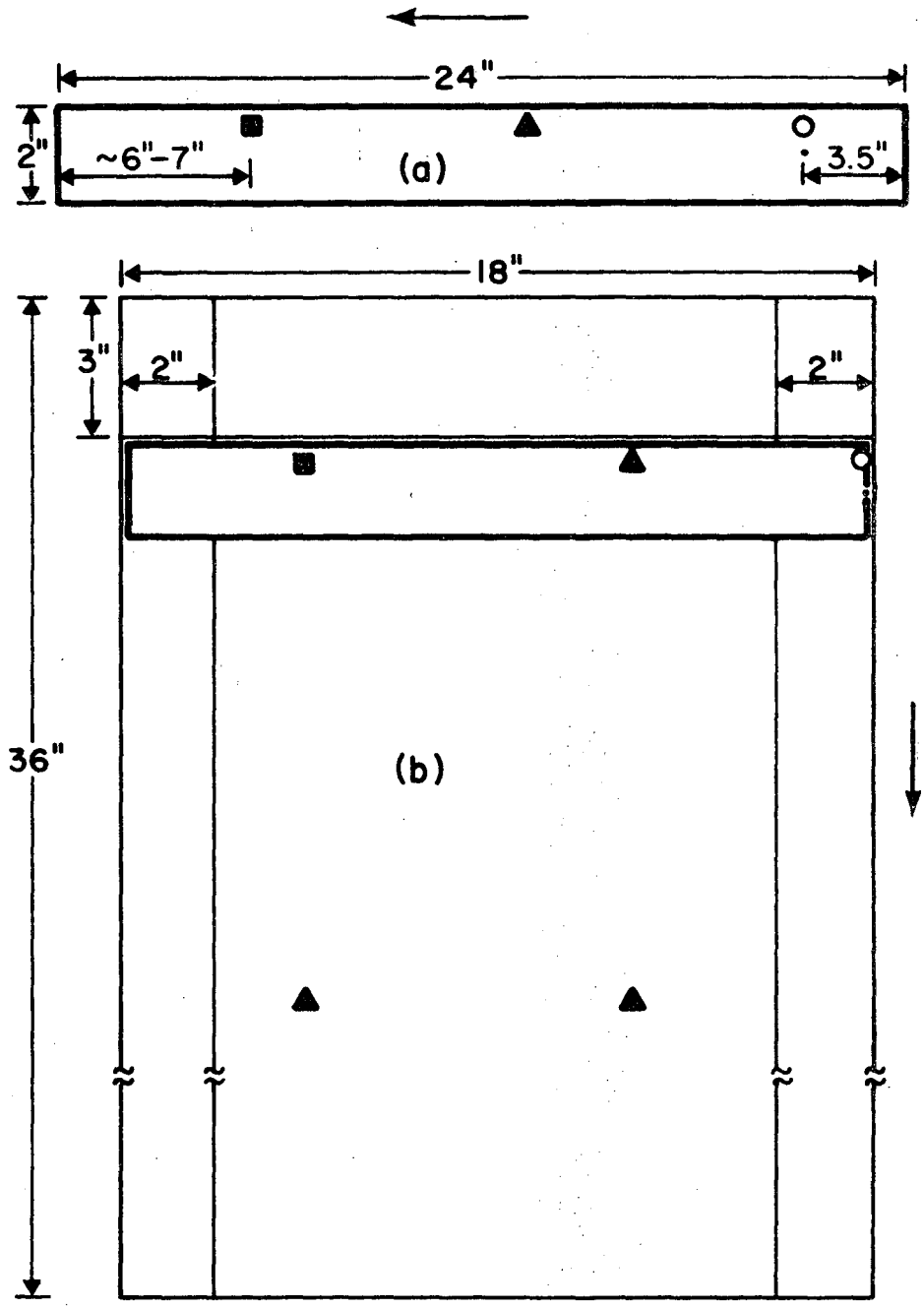
(iv) Spotting and the first dimensional ionophoresis:

A cellulose acetate strip of 2"x24" was cut out from a roll of 2"x10'. It was first wet with the pH 3.5/7 M urea buffer. To avoid inclusion of air bubbles wetting was done by floating the strip on the buffer in a petri dish. Excess buffer should be blotted. The wet strip was placed on a rack and lowered into the tank containing the same buffer. Pre-ionophoresis was carried out at 2,500 volts for 30 minutes. At the end of this equilibration, the rack was taken out of the tank, and the point of application, which was 3.5" from the cathode end of the strip, was blotted. The sample was immediately applied as a spot and allowed to soak in while the strip was still on the rack. Two spots of the colored marker were applied one on each side of the sample. Several precautions one must note before spotting:

(a) Avoid overloading the cellulose acetate strip; this will result in bad streaking in the first dimension. Best results were obtained with less than 100 μg of the RNA digest.

(b) The size of the spot should be kept very small in

0 0 0 0 3 8 0 0 2 9 1



XBL 7110-1574

Fig. 3. Lay out of ionophoresis papers

(a) Cellulose acetate strip (2"x24"),
(b) DEAE-paper (18"x36");
• Point of application of sample (origin), ○ samples of dye mixture, ■ red dye at the end of run, ▲ blue dye at the end of run, ⟶ ionophoresis directions.

order to achieve good resolution. Thus all spottings were carried out in drawn out melting point tubes. Each sample was spotted at short intervals.

(c) Care must be taken to avoid having the strip dry out while the sample is being applied. 5-10 μl of sample was found most convenient as it rapidly soaked into the strip.

After the sample was completely soaked in, the rack was immediately lowered into the same tank. Ionophoresis was run at 2,500 volts until the red marker had traveled 15" from the origin. This could be done in about 3 hours. The majority of the oligonucleotides was found to migrate slower than the red marker and faster than the blue.

(v) Transferring and the second dimensional ionophoresis:

Fractionation in the second dimension was carried out on sheets of DEAE-paper. The material on the cellulose acetate strip was transferred onto the DEAE-paper by the following blotting procedure. A sheet of DEAE-paper 18"x36" was laid on a glass plate. The cellulose acetate strip on which the oligonucleotides had been fractionated was allowed to drip off the excess Varsol. Before the buffer dried the strip was laid on the DEAE-paper about 3" from the cathode end. A pad of three strips of Whatman 3 MM paper 2"x18" that had been soaked in water was then placed on top of the strip and a glass plate placed on top to press the strips together

evenly. Water from the pad passed through the cellulose acetate strip carrying the oligonucleotides with it and onto the DEAE-paper. Being acidic they were held on the DEAE-paper by ion-exchange and remained in the position at which they were first washed on. To ensure complete transfer, more water was added to the pad while it is still in position. The water was allowed to soak through until a 4" to 5" wide strip of the DEAE-paper had been wet. If this was done carefully more than 90% of the material was transferred.

Since urea would cause streaking in the second dimension, it should be removed by washing the wet area of the DEAE-paper twice in 95% ethanol, each time with fresh ethanol, then with 100% ethanol. The DEAE-paper was allowed to dry while mounted on a rack.

Second dimensional ionophresis was performed at ninety degree to the direction of that on cellulose acetate strip. The DEAE-paper was wet with 7% formic acid pH 1.9 while still mounted on the rack. Wetting was started on both sides of the line, where the oligonucleotides were transferred, with the rack laid on the side. The fronts of the solution were allowed to meet along this line. Then the rest of the paper was wet. It was advisable not to overwet the paper. Wetting could be done evenly with a spray or a wash bottle. The wet DEAE-paper on the rack was lowered into the iono-

phoresis tank that contained 7% formic acid with the oligonucleotides near the cathode. Ionophoresis was carried out at 1,250 volts for 16 hours or until the blue marker had traveled 14" from the origin. The DEAE-paper was finally dried in air while still on the rack.

(vi) Radioautography:

This two-dimensional high-voltage ionophoresis fractionation is referred to as the fingerprinting procedure. The resultant radioautographs are called the fingerprints. To prepare radioautographs, the DEAE-paper was marked with radioactive ink (^{14}C) and cut to convenient sizes for exposure. The DEAE-paper containing the fractionated oligonucleotides was put in a light proof X-ray exposure holder with one or two sheets of no-screen X-ray film. The desired time for exposure to obtain a satisfactory fingerprint was summarized in Table III. Films were developed after sufficient time was given for exposure.

G. Structure of the Oligonucleotides

(i) Elution:

The positions of the oligonucleotides on the DEAE-paper were determined from the radioautograph by matching the ^{14}C ink marks on the film with that on the paper. The spots were cut out to a point as shown in Fig. 4 (a). The oligonucleotides were eluted with 2 M triethylamine bicarbonate ($TEA\text{-}HCO_3^-$) at pH 8. This was prepared by

0 0 0 0 3 8 0 0 4 9 3

Table III

Radioautogram Exposure Time

Amount of activity (CPM)	Exposure time
60,000 or less	2-3 weeks
300,000	1-2 days
500,000	16-30 hours
1,000,000	8-16 hours
2,000,000 and more	3-5 hours or less

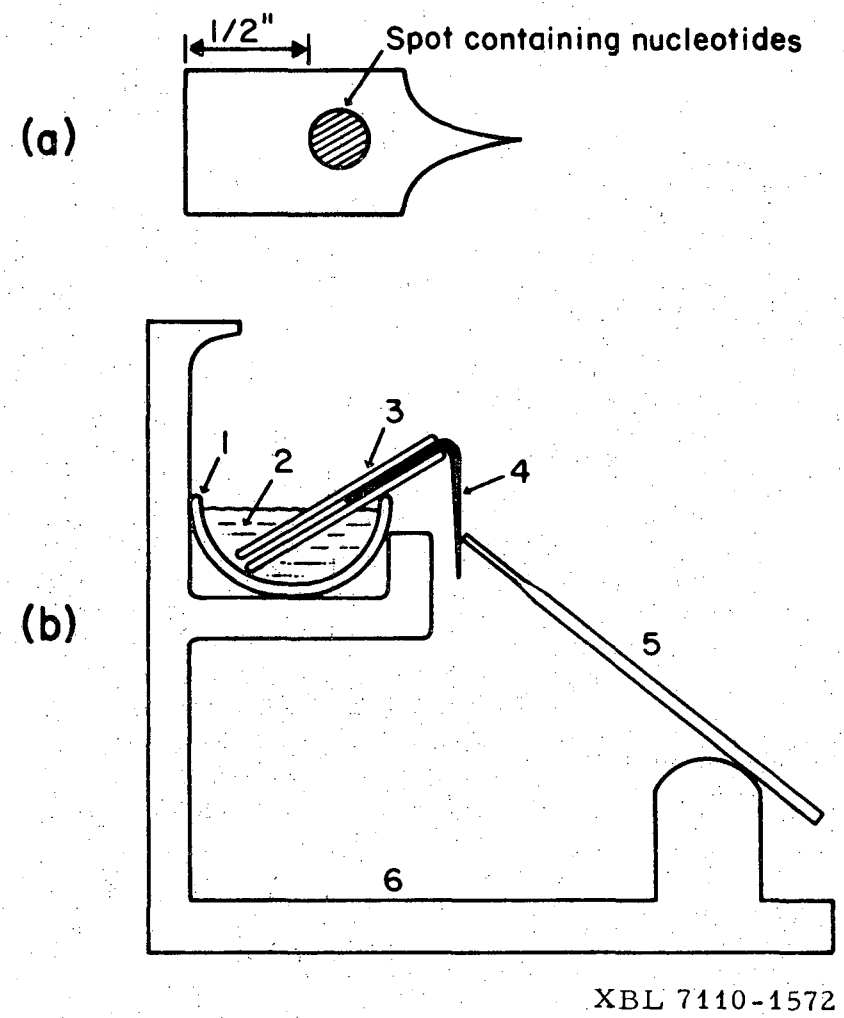


Fig. 4. Elution set-up

(a) Cut out,
(b) Cross section of the set-up:
(1) glass trough,
(2) 2 M TEA-HCO_3^- pH 8,
(3) microscope slides 1"x3",
(4) DEAE-paper cut out as shown in (a),
(5) drawn out melting point tube,
(6) supports for the trough and tube.

passing CO_2 into a mixture of 70 ml of water and 30 ml of redistilled triethylamine (TEA) until all TEA had dissolved into water as TEA-HCO_3^-. Elution was carried out as shown in Fig. 4 (b). It was very rapid by this method and was complete within half hour. For each spot not more than 100 µl was usually collected in the drawn out melting point tube.

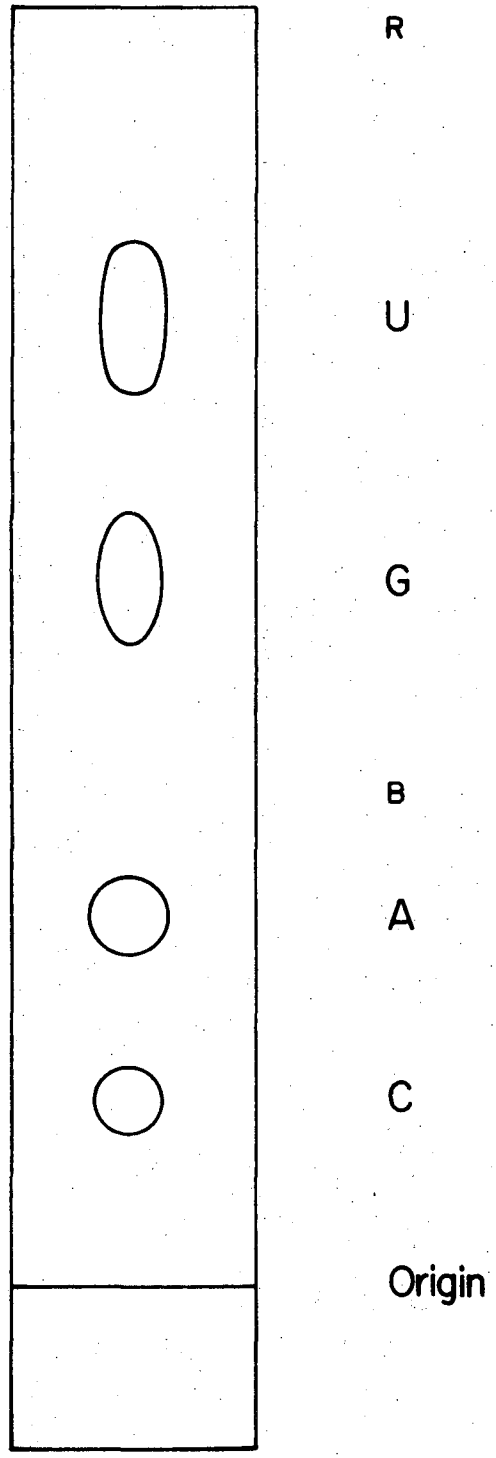
The eluted materials were then transferred onto small petri dishes and dried in a partial vacuum. If a good vacuum was used, TEA-HCO_3^- would bubble badly. The residues were redissolved in 10 µl of water and taken to dryness three times to ensure complete removal of the TEA-HCO_3^-. The residual oligonucleotides were finally taken up in the desired buffer for structural analysis described in the following sections.

(ii) Base composition:

To determine the base composition of an oligonucleotide 10 µl of 0.2 N KOH was used to dissolve each residue. The KOH solution containing the oligonucleotides was drawn into a drawn out melting point tube by capillary action. The end was sealed off. Hydrolysis was carried out at 37° C for 16 hours. After incubation the digest was applied directly to Whatman No. 1 paper (18"x23") for separation. Each digest was applied as a spot 3.5 inches from one end of paper with 1.5 inch spacing. A total of 10 to 12 digests could be run at one time. Two

dye spots were also applied along two sides. Then the paper was wet with pH 3.5 buffer (no urea). Care must be taken not to smear the spots during the wetting. Ionophoresis was run in the same buffer at 2,000 volts for 2 hours or until the red marker had reached the anode end of the paper. The paper was allowed to dry in air while on the rack. The four mononucleotides were well separated by this system (Fig. 5) with C and A residues traveled slower than the blue spot and G and U residues traveled faster than the blue but slower than the red spot.

The position of the four mononucleotides can be determined from its radioautograph as shown in Fig. 7. They can also be detected by UV light. In this case, along with each KOH digest, 5 to 10 μl of a mixture containing 1 mg/ml each of the unlabeled mononucleotides was also spotted. The idea was to use these unlabeled mononucleotides as carriers which can be detected by a UV light source. However, pyridine also absorbes in the UV region so that it must be removed by washing with ethylacetate. The spots were then cut out and the activity of each mononucleotide was directly estimated on the scintillation counter. For products from T_1-RNase digests, a G residue only occurs once at the 3' end of each oligonucleotide. Thus the ratio of each other bases to the one G residue gives the base composition as well as the nucleotide chainlength. In the case of PAN-RNase digests,



XBL 7111-1703

Fig. 5. Diagram showing the position of mononucleotides from KOH digestion. Ionophoresis was done at pH 3.5 on Whatman No. 1 paper (2,000 volts, 2 hours). R and B represent the red and blue markers respectively.

a C or U residue occurs only once in each oligonucleotide, thus comparisons were made with respect to them.

(iii) Sequence analysis:

To determine the base sequence of the oligonucleotides of an enzyme digest it is best to treat with a second enzyme. For T_1-RNase digests, the oligonucleotides were further degraded with PAN-RNase. T_1-RNase cleaves the phosphodiester bond following a G residue, whereas PAN-RNase is specific for pyrimidine nucleotides. If products of T_1-RNase digests were treated with PAN-RNase, one would obtain a group of oligonucleotides consisting A_nG, A_nC and A_nU, where n may have values of 0, 1, 2, 3, 4 or higher. These oligonucleotides could be separated by ionophoresis at pH 3.5 on DEAE-paper.

Secondary enzyme digestion was carried out in 0.001 M EDTA- 0.01 M Tris (EDTA-Tris) buffer pH 7.4 containing 0.2 mg/ml PAN-RNase. The oligonucleotide residues of T_1-RNase digest eluted as described in section G(i) were each dissolved in 10-15 µl of this buffer and taken up in a drawn-out melting point tube. The mixtures were incubated at 37° C for 4 hours with ends sealed. At the end of the incubation, the sealed ends were broken, each mixture was squirted out and mixed well with 2-3 µl of 0.5 N HCl, taken up again in the same capillary tube, ends resealed and incubated for an hour at 37° C. The treatment with 0.1 N HCl was to break down any cyclic phosphates.

The digests were then applied directly from the capillary tubes onto the DEAE-paper (18"x36"). They were applied as 0.5 inch lines one inch from each other. 3.5 inch spacing from one end of the paper should also be allowed. Ionophoresis was run at pH 3.5 (no urea) and 1,000-1,500 volts until the blue marker had moved 9 inches from the origin. The DEAE-paper was wet with buffer before lowering into the ionophoresis tank that contained the same buffer. Wetting was as described in section F(v). At the end of the ionophoresis, the paper was allowed to dry while still on the rack.

A small sample (30-40 μg) of ^{32}P-TMV-RNA in EDTA-Tris buffer pH 7.4 with an activity of 2×10^6 CPM was digested with 2 enzyme units of T_1-RNase and 2 μg of PAN-RNase at 37° C for 4 hours. After 0.1 N HCl treatment the products were fractionated on DEAE-paper by ionophoresis at pH 3.5 as described in the previous paragraph. Fig. 6 shows the resultant radioautograph and its diagram. The fractionated products were further identified from their base composition. They were eluted with TEA-HCO_3^- as in G (i) and hydrolysed with alkali as in G(ii). The resulting mononucleotides were fractionated by ionophoresis on paper at pH 3.5, located by radioautography (Fig. 7), and measured by scintillation counting. The counting results were tabulated in Table IV. Since each oligonucleotide must terminate at the 3' end by a

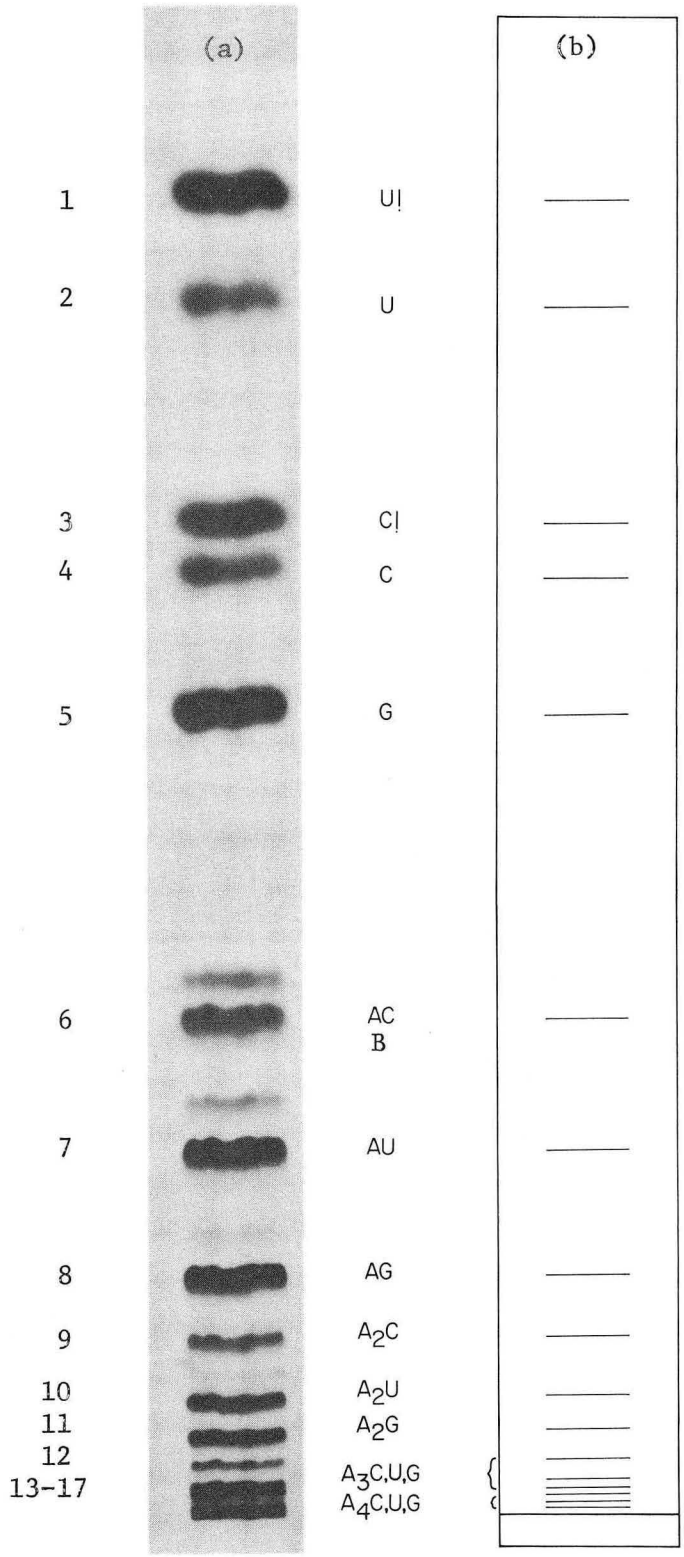


Fig. 6. (a) Radioautograph showing the position of products from PAN-RNase digests of oligonucleotides from T_1-RNase digests on ionophoresis at pH 3.5. B shows the position of the blue marker.
(b) Its diagram.

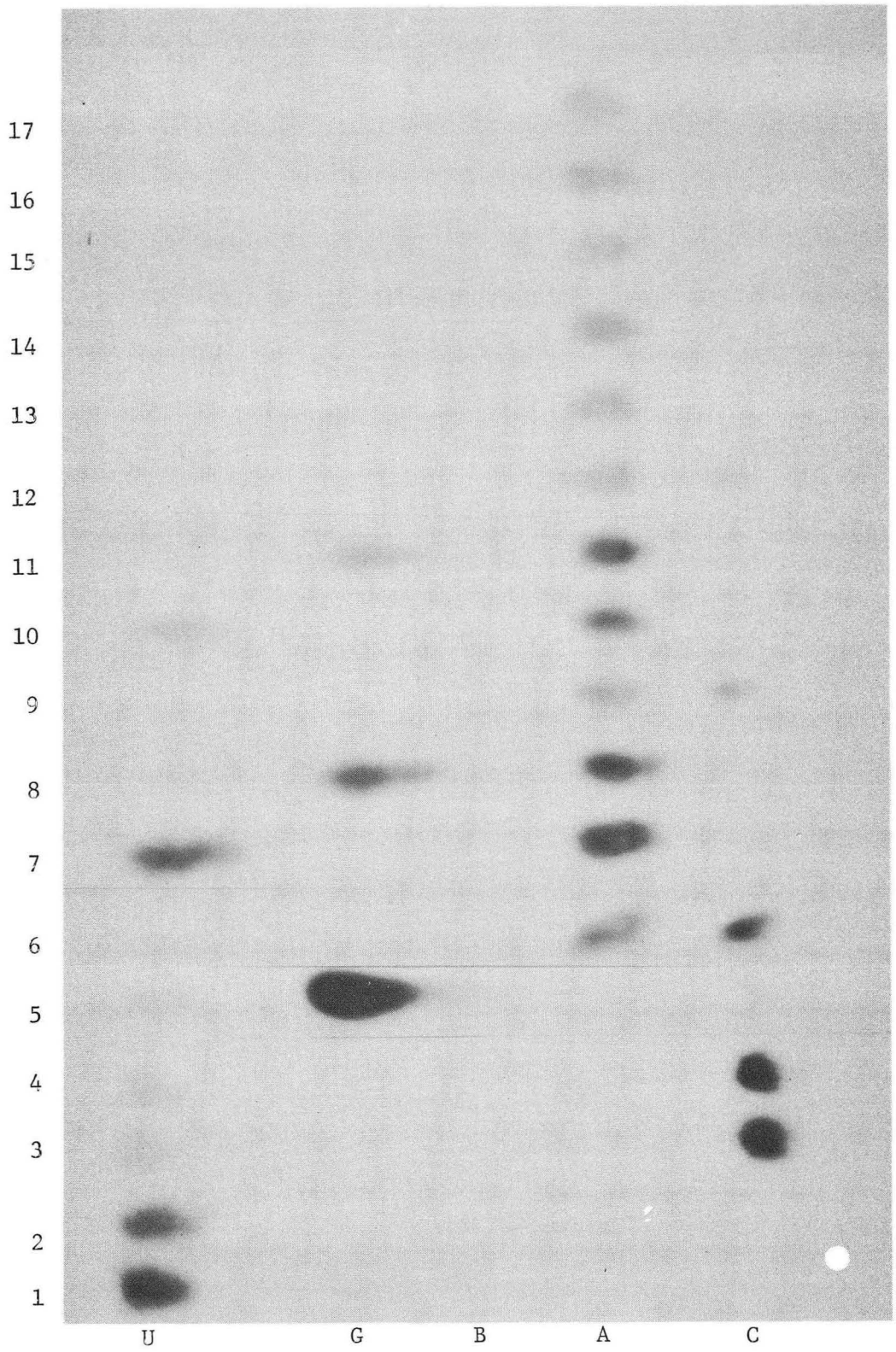


Fig. 7. Radioautograph showing the position of mononucleotides from KOH hydrolysis of PAN-RNase digests. B is the position of the blue marker. Numbers correspond to those in Fig. 6.

Table IV

Nucleotide Base Composition

Spot Number (Fig. 6)	Base Composition				Structure Deduced
	A	C	U	G	
1	-	-	x	-	U!
2	-	-	x	-	U
3	-	x	-	-	C!
4	-	x	-	-	C
5	-	-	-	x	G
6	0.98	1.00	-	-	AC
7	1.01	-	1.00	-	AU
8	0.96	-	-	1.00	AG
9	1.92	1.00	-	-	A_2C
10	1.98	-	1.00	-	A_2U
11	2.05	-	-	1.00	A_2G
12	2.91	1.00	-	-	A_3C
13	3.15	-	1.00	-	A_3U
14	3.25	-	-	1.00	A_3G
15	3.79	1.00	-	-	A_4C
16	3.71	-	1.00	-	A_4U
17	4.27	-	-	1.00	A_4G

G, C or U residue, the base compositions were determined with respect to one of these residues. The correct structures with various numbers of A residues are given in the last column. This was used as a standard for determining the structures of oligonucleotides fractionated by this system.

0 0 0 0 3 8 0 0 2 9 7

III. RESULTS

A. T_1-RNase Fingerprint of Whole ^{32}P-TMV-RNA

From each preparation of ^{32}P-TMV, 5 x 10^6 CPM of material was taken. Its RNA was extracted by phenol. The resultant whole ^{32}P-TMV-RNA was completely hydrolysed with 2 enzyme units/A_{260nm} of T_1-RNase at 40° C for 4 hours. 300,000-500,000 CPM of the products was fractionated by the two-dimensional high-voltage ionophoresis as described in section II F. A total of six fractionations have been carried out with RNA extracted from five different virus preparations. The results were completely reproducible. Fig. 8 shows one of the two-dimensional fingerprints of the digestion products. Although resolution was good no attempts were made to identify the materials present in each spot. Careful base sequence analysis has been done by Sanger and his colleagues on T_1-RNase digested ^{32}P-labeled 16S and 23S ribosomal RNA (rRNA) (Sanger *et al.*, 1965). The general fingerprint pattern of T_1-RNase digested ^{32}P-TMV-RNA looks very similar to both 16S and 23S rRNA, but only more complicated in the larger fragment region. TMV-RNA is almost four times as long as 16S and twice as long as 23S rRNA. It is expected to have more larger fragments in complete T_1-RNase digests. The following is to discuss some of the characteristics of such a fingerprint.

Since only uracil among the four bases carries no positive charges in pH 1.9 or 3.5, its ribophosphate will have the highest negative charges in those pH's. Thus when fractionated by ionophoresis, any fragments containing U residues will move faster by

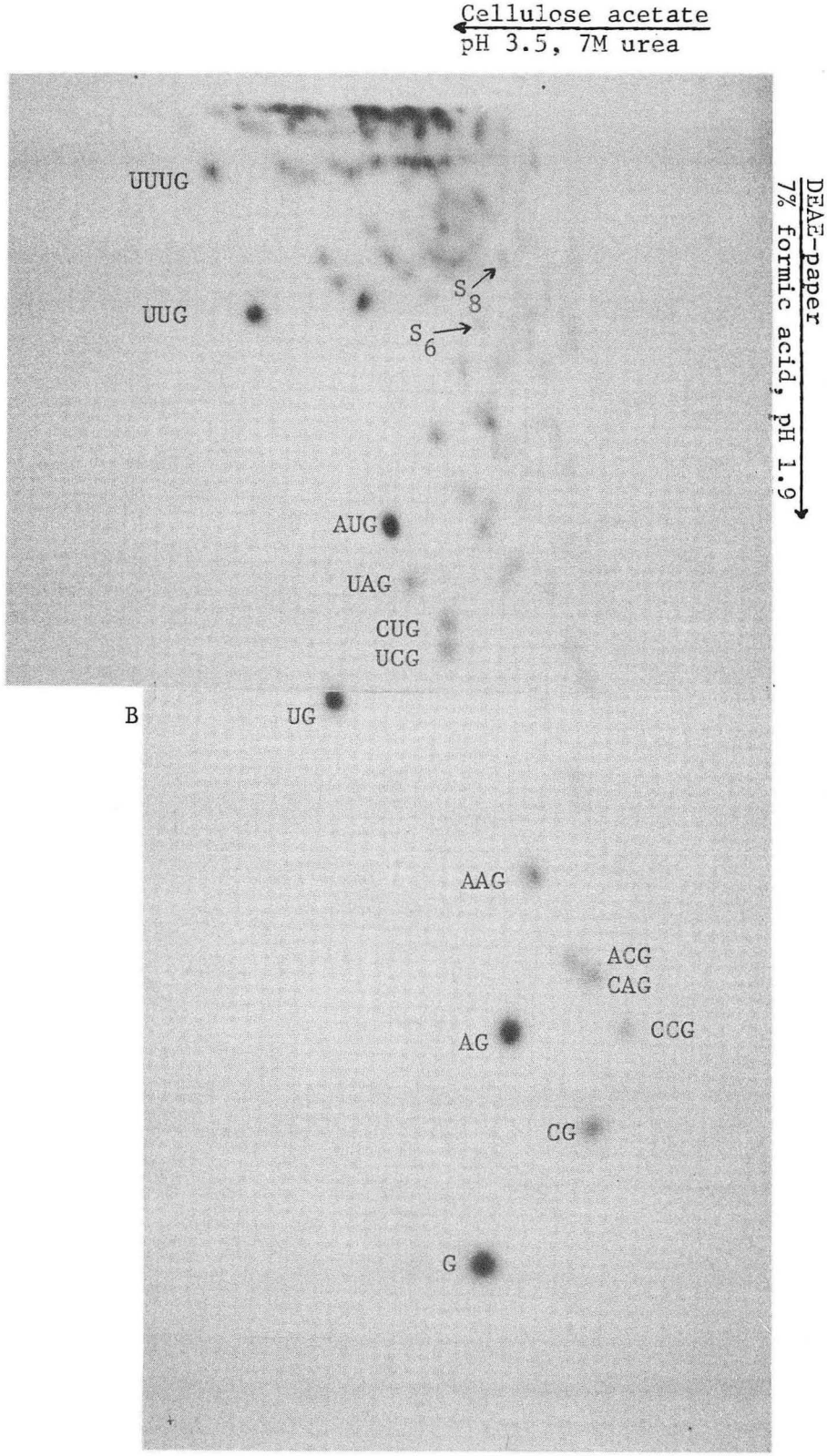


Fig. 8. Radioautograph of a complete T_1-RNase digest of ^{32}P-TMV-RNA on ionophoresis on cellulose acetate strip at pH 3.5 (2,500 volts, 3 hours) as the first dimension and on DEAE-paper at pH 1.9 (1,250 volts, 16 hours) as the second dimension. B is the position of the blue marker on the second dimension.

0 0 0 0 3 8 0 0 2 9 8

ionophoresis but slower by ion-exchange than those with no U residues. The more U residues a fragment contains the bigger the difference will be. As a matter of fact the movement of each fragment is essentially governed by the number of U residues it contains. This is evident in the fingerprint shown in Fig. 8. The two-dimensional ionophoresis procedure fractionates complete T_1-RNase digests of ^{32}P-TMV-RNA into groups. All fragments in the same group contain the same number of U residues. The group containing no U residues travels slowest on cellulose acetate strip because electrophoretic effect is the only force that moves the nucleotides. On the other hand it travels fastest on the DEAE-paper where ion-exchange opposes the ionophoresis effect. Therefore at the end of the fractionation this group occurs in the lower right hand portion of the fingerprint. It is represented by mononucleotide G in Fig. 8. By the same reasoning, the group containing one U residue is the next upper left group as designated by dinucleotide UG. And next, the UUG group with two U residues. Still next, the UUUG group with three U residues, and etc.

In each group the upper boundary contains oligonucleotides with increasing numbers of A residues and the lower boundary contains oligonucleotides with increasing numbers of C residues. Those in between are oligonucleotides containing various numbers of A's and C's. Nucleotides G, AG, CG, UG, and the nine trinucleotides are shown in Fig. 8 to demonstrate these trends. (These sequences will be confirmed in the next section.) Lines can be drawn connecting all oligonucleotides with increasing numbers of A's or

C's. By extending these lines, one would be able to estimate the base composition of certain longer oligonucleotides. What is more useful is that one can draw lines connecting all di-, tri-, or tetranucleotides and etc. This enables one to estimate the chain-length of an oligonucleotide present in a spot from its position in the fingerprint. This was also used to locate a certain hexanucleotide which was used as a reference for the chainlength determination of ^{32}P-TMV-RNA.

B. The Chainlength Determination of ^{32}P-TMV-RNA

The position of oligonucleotides on the DEAE-paper, on which the T_1-RNase digests of whole ^{32}P-TMV-RNA were fractionated, was determined as described in section II G(i). Each spot was cut out and its radioactivity in counts per minute (CPM_i) was counted in a scintillation counter. The total chainlength (L) of the whole ^{32}P-TMV-RNA molecule could be calculated from

$$L = \frac{\sum_i CPM_i}{Act(P)} \qquad (1)$$

provided the radioactivity of one phosphate (Act(P)) was known. To determine the radioactivity of a single phosphate, hexanucleotide S_6 (Fig. 8) was chosen as follows. In the group of oligonucleotides represented by UG, lines were drawn connecting all trinucleotides, tetranucleotides, etc. By doing so all hexanucleotides in this group were located. Spot S_6 appeared to be uniquely resolved and was chosen as a reference. As a rough theoretical approximation, a particular hexanucleotide can be obtained only once from a random

polymer of 4,000 residues long. Thus it was justified to assume that spot S_6 contains only one hexanucleotide and it only occurs once along the whole ^{32}P-TMV-RNA chain, then

$$\text{the apparent Act(P)} = \frac{CPM_{S_6}}{6} . \qquad (2)$$

This calculation has been performed on all six fingerprints of T_1-RNase digested whole ^{32}P-TMV-RNA. The activity of one phosphate was determined separately for each case. The chainlengths obtained were of the order of three thousand nucleotides, which is about half of the commonly accepted value (Fraenkel-Conrat, Molecular Basis of Virology, 1968). This suggested spot S_6 contained two sequence isomers of six nucleotides which were not separable by this fractionation procedure. Thus an octanucleotide S_8 (Fig. 8) was chosen and the activity of one phosphate was calculated as for S_6. In each case the value obtain from S_8 was half of that from S_6. Here theoretical estimation for the chance of a particular octanucleotide to occur in a random polymer was one in every 4^8 (or 65,000) nucleotides. Therefore the actual activity of each phosphate would be equal to the CPM of S_6 divided by 2 x 6,

$$\text{Act(P)} = \frac{CPM_{S_6}}{2 \times 6} . \qquad (3)$$

The total chainlength of ^{32}P-TMV-RNA was determined based on this number as a reference.

Table V lists the chainlength obtained for ^{32}P-TMV-RNA in column 4. Columns 2 and 3 are the values for the activity of one phosphate obtained from S_6 and S_8 respectively. All chainlength

Table V

Chainlength of Whole ^{32}P-TMV-RNA

^{32}p-TMV preparation	$\frac{CPM_{S_6}}{2 \times 6}$	$\frac{CPM_{S_8}}{8}$	$\frac{\sum_i CPM_i}{Act(P)}$	$\sum_i l_i$
1	40.7	39.4	6,180	5,855
2	89.0	-	6,110	-
3	25.7	-	6,540	-
4	19.0	18.4	6,250	6,057
5a	39.6	39.0	6,265	6,120
5b	73.0	69.5	6,360	6,129
Average			6,284 ± 2.4%	

values agreed with each other within ± 2.4%. This suggested that the two-dimensional fingerprinting procedure was completely reproducible. The small discrepancy was probably due to (1) the somewhat incomplete transferring of large oligonucleotides from the cellulose acetate strip onto the DEAE-paper; (2) inclusion of some short virus rods either broken or incompletely synthesized in the virus preparation. The missing of some large oligonucleotides due to incomplete transferring was responsible for obtaining chainlength shorter than the true value. If the contaminating short virus rods contained the hexanucleotide used as a reference, then the result would also be on the short side. On the other hand, inclusion of short rods not containing the reference hexanucleotide would give rise to a number larger than the true value. However, the average chainlength of ^{32}P-TMV-RNA determined by this method agreed within 1% of those values established by other biochemical or physical chemical methods (Klug & Caspar, 1960; Caspar, 1963).

Since the activity of one phosphate is a fixed number it can be put into the summation sign in equation (1) and rearranged mathematically as

$$L = \frac{\sum_i CPM_i}{Act(P)} = \sum_i \frac{CPM_i}{Act(P)} = \sum_i l_i \ . \qquad (4)$$

$l_i = \frac{CPM_i}{Act(P)}$ can be considered as the effective chainlength contributed by all fragments in spot i. Summing l_i over all spots will give the total chainlength of ^{32}P-TMV-RNA. The results obtained in this way are summarized in the last column in Table V. Unfortunately,

two of the fingerprints were not perfectly resolved for some larger fragments, therefore only four numbers are listed. The comparatively shorter chainlength obtained in this way was due to round-off errors when computing l_i.

Although this was a more elaborate way of getting the total chainlength, an obvious reason for undertaking such a task was to make sure whether the activity of one phosphate was really chosen correctly. The chainlength of the oligonucleotides in spot i was estimated from its position in the fingerprint. Dividing l_i by this chainlength would give the molar ratio of these oligonucleotides with respect to one ^{32}P-TMV-RNA chain. If the value for Act(P) was chosen correctly any molar ratio should be an integer greater or equal to unity. This was indeed the case with Act(P) obtained from equation (3). With Act(P) values calculated from equation (2) molar ratio of half integers was obtained for many oligonucleotides and molar ratio of 1/2 was obtained for some large oligonucleotides. It was then concluded from these evidences that equation (3) was correct. Therefore a chainlength of 6,284 is reported here for the whole ^{32}P-TMV-RNA with ± 2.4% uncertainty.

C. SDS Polar Stripping of Protein Subunits from ^{32}P-TMV

(i) Stripping Condition:

The behavior of the SDS stripping of protein subunits from TMV was first studied by Hart (1958). He discovered that treatment of TMV rods with SDS at 85° C tended to remove protein subunits from only one end of the rod. This end was later determined by May & Knight (1965). They used the two specific exoribonucleases to

hydrolyse the end of the RNA uncovered during the SDS stripping at 37° C. The infectivities of the enzyme treated PSV were studied. It was found that treatment with spleen phosphodiesterase did not alter the infectivity of the PSV. However, loss of infectivity, when the PSV was treated with snake venom phosphodiesterase, was found parallel with those of TMV-RNA when treated similarly. Snake venom phosphodiesterase has been shown to attack polynucleotides from the 3' end (Singer et al., 1958), whereas spleen phosphodiesterase works from the opposite end (Razzell & Khorana, 1961). Therefore, it was concluded by May & Knight that the end of the RNA exposed by SDS is the 3' end and the protein subunits were removed sequentially from this end of the viral rods. This kind of progressive removal of protein subunits from the 3' end of TMV will be referred to as the SDS polar stripping throughout this work.

A large number of work has been reported based on this SDS polar stripping property of TMV. The local lesion gene was located at a position approximately one quarter of the length of the RNA starting from the 5' end (Kado & Knight, 1966). The coat protein gene of TMV was located similarly in the first half of the RNA starting from the 5' end (Kado & Knight, 1968). Mandeles (1968) has reported the locations of three unique sequences, found in the complete T_1-RNase digest, in TMV-RNA. Ohno et al. (1971) were able to isolate fragments of TMV-RNA containing exclusively the 5' end and carried out their reconstitution experiments, which eventually led to the conclusion of polar reconstitution behavior of TMV.

Kado & Knight carried out their polar stripping reaction with 1% SDS in 2% solution of TMV at 37° C for 1 1/2 - 3 hours. They used this relatively vigorous condition to achieve an extent of stripping as high as 50-80%. However the prolonged reaction resulted in a wide distribution of the extent of stripping, so that they had to fractionate the PSV by sucrose density gradient centrifugation. In this regard, Symington & Commoner (1967) have described stripping experiments carried out with 0.02% SDS in 100-200 μg/ml (0.01-0.02%) solution of TMV at 62°-68° C. Much lower concentration in both TMV and detergent was used. But a higher temperature was employed in order to achieve large extent of stripping in a relatively short time (5-15 minutes). Their results showed an improved distribution of PSV. The foregoing results indicated the rate of SDS stripping of protein subunits from TMV was greatly enhanced by high temperature and reduced by low SDS concentration. The work reported here was devoted to obtain stripping as little as 1-5%. The stripping condition used was chosen carefully so that the desired small extent of stripping can be obtained within a reasonably short time.

A series of stripping experiments were carried out at room temperature with SDS concentrations varied over a range of 0-1.5% meanwhile keeping other conditions constant. The extent of stripping was followed by observing the scattering at 400 nm as described in section II E. As a result, 0.5% of SDS was found to produce up to 5% stripping within 25 minutes (Fig. 9). Reaction proceeded too slowly in less than 0.5% of SDS; this was likely to produce a wide distribution of PSV. With an SDS concentration higher than 0.5%

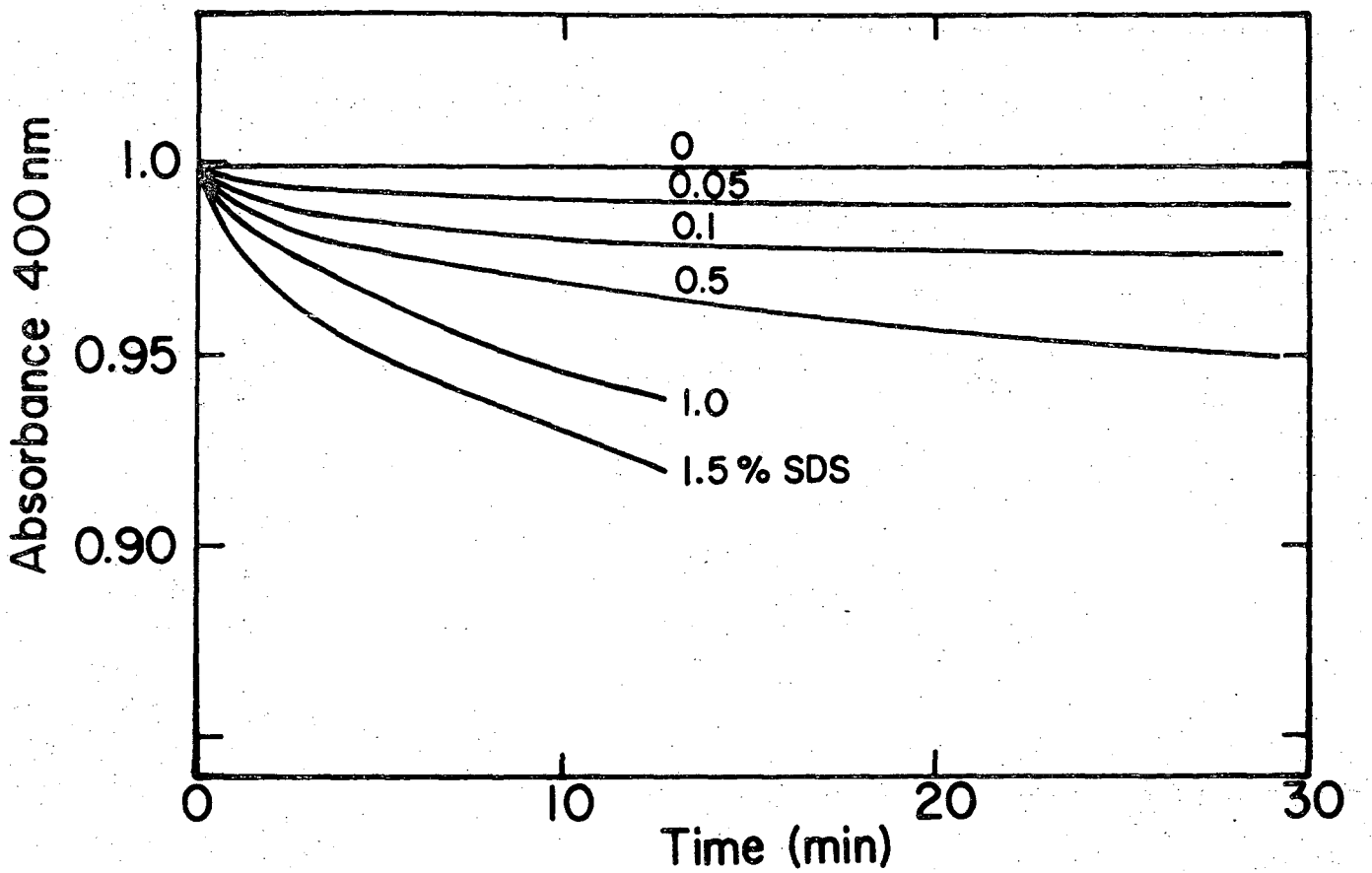


Fig. 9. SDS stripping of TMV as followed by the scattering at 400 nm.

the initial stripping rate was too rapid to control accurately. All these experiments were done with 4-8 mg/ml of unlabeled TMV in EDTA-cac buffer pH 7.2 containing 20-40 μg/ml of bentonite. The use of bentonite is to prevent degradation of the exposed RNA by extraneous ribonucleases. Singer & Fraenkel-Conrat (1961) have reported that a weight ratio of bentonite to enzyme of 1.5 or more completely inhibits the action of PAN-RNase and other plant nucleases. Thus an upper estimate of 10-20 μg/ml of such RNases as impurity in the buffer calls for 20-40 μg/ml of bentonite. Furthermore, varying concentration of TMV within 4-8 mg/ml does not affect the rate of stripping. The absorbance of this solution at 400 nm is between 0.5-1.0 which is the most reliable region to work with. Therefore the stripping conditions used in this work are:

^{32}P-TMV, 4-8 mg/ml;
buffer, EDTA-cac pH 7.2;
bentonite, 20-40 μg/ml;
temperature, room temperature; and
SDS, 0.5%.

(ii) The extent of stripping:

Using the conditions described previously, attempts were made to strip off 1, 3, and 5% of the coat protein subunits of ^{32}P-TMV. Table VI summarizes the results obtained from these three independent stripping experiments. In each experiment the ^{32}P-labeled PSV was obtained and purified as described in II E. The resultant ^{32}P-PSV free of bentonite was obtained in 1-2 ml of 0.001 M Tris buffer at pH 7.4. Three 10 μl aliquots were taken

Table VI

SDS Polar Stripping of Protein Subunits of ^{32}P-TMV

	1	2	3
Total CPM of ^{32}P-TMV	20×10^6	45×10^6	20×10^6
A^i_{400nm}	0.580	0.975	0.605
A^f_{400nm}	0.575	0.945	0.575
Time (min) of Stripping	1-2	10-12	20-25
Volume (ml) of ^{32}P-PSV	1.3	2.0	1.2
CPM/10 µl of ^{32}P-PSV	98,450	171,828	129,000
CPM/10 µl of T_1-RNase digest	1,145 1,169c	5,288 5,390c	5,800 5,920c
% Stripping	1.19	3.14	4.58

A^i_{400nm} is the absorbance measured at 400 nm at the time of adding SDS, A^f_{400nm} is the absorbance measured at 400 nm at the time the stripping was quenched, and c stands for correction due to radiation decay.

and their radioactivities counted in a scintillation counter. The average activity was listed in Table VI for each extent of stripping. The exposed 3' end of the RNA was hydrolysed by T_1-RNase stored at 1,000 enzyme units/ml. Two enzyme units of T_1-RNase were added to the ^{32}P-PSV solution and the mixture was incubated at 40° C for 2 hours, then another enzyme unit of T_1-RNase was added to the mixture and incubation continued for another 2 hours at 40° C. Hydrolysis was stopped by adding 100 μg of bentonite stored at 55 mg/ml. This mixture was centrifuged twice at 37,000 rpm in the 40 rotor of model L centrifuge for 3 hours. The resultant supernatant contained the T_1-RNase digests of the 3' terminus of ^{32}P-TMV-RNA. Three 10 μl aliquots of this supernatant were taken and counted. The average radioactivity for each extent of stripping was listed in Table VI. The ratio of the ^{32}P activity of the T_1-RNase digest of the 3' terminus of the RNA to that of the untreated ^{32}P-PSV gives a direct estimate of the extent of stripping. However, the following two corrections must be made for calculating the exact extent of stripping. ^{32}P decays with a half life of only 14.3 days as e^{-kt}, where $k = \frac{\ln 2}{14.3}$ and t is time. Knowing the time it takes to obtain the T_1-RNase digests from the PSV, corrections can be carried out most straight forwardly. (This correction due to radiation decay is done in this work whenever necessary.) Secondly, the change of volume going from PSV to the final T_1-RNase digest should also be considered. However, as only a total of 5-6 μl of T_1-RNase and bentonite was added to 1-2 ml of the PSV solution, it was usually negligible compared to the radiation decay of ^{32}P. (One day takes

away 5% of the activity!) The corrected values for the activity of the T_1-RNase digests are also listed in Table VI. The percent strippings were calculated based on them.

This is a far more convenient and accurate method of determining the exact extent of stripping than the conventional method which depends solely on how accurately the amount of the stripped protein is measured. Since 1 mg/ml of TMV protein has an absorbance of only 1.27 at 282 nm (Fraenkel-Conrat, 1957), the concentration at which the absorbance reading is reliable should be at least 0.25 mg/ml. However in this work, SDS polar stripping was not only carried out with a few milligrams of ^{32}P-TMV, the stripped protein was also obtained in a large amount of buffer. (The stripping was stopped by diluting with more than 10x of cold buffer.) Thus an estimate of the amount of protein of as little as 1-5% of a few milligrams in a volume of 20-30 ml, i.e. less than 10 μg/ml, can hardly be accurate even by use of the very sensitive Lowry's method (Lowry _et al_., 1951). Besides, one is also required to measure all volumes involved precisely. On the contrary, as long as the volume of the reaction mixture was kept unchanged, one only needs to measure the ^{32}P activity of 10 μl each from the ^{32}P-PSV and its T_1-RNase digest. Their ratio, regardless of the total volume, gives the precise extent of stripping.

(iii) The distribution of stripping:

It is important to keep in mind that the percent stripping listed in Table VI is the average degree of stripping. SDS strips protein subunits sequentially from the 3' end of TMV, but it does

not remove the same subunits in all viral rods at the same time. It would be ideal if the SDS stripping could be stopped at the same position in all viral rods, but on the contrary, some kind of distribution has been obtained. The precision of the sequence analysis presented in later sections of this report depends largely on this distribution in chainlengths of the shortened rods obtained for any extent of stripping. Therefore it was attempted to investigate the distribution of the stripping products obtained in this work.

A number of physical chemical methods have been used to study the rod length of TMV (Boedtker & Simmons, 1957). Perhaps the most direct and widely used method of measuring the rod length is the electron microscopy. It has been used to determined the length of the whole virus rod (William & Steere, 1951; Hall, 1957). More interestingly Symington & Commoner (1967) have investigated the rod length of SDS treated TMV by using this technique. They have noticed a small distribution even in the untreated whole TMV and this distribution widens as the rods become shorter. However, the accuracy of differentiating the distributions of 1% stripped TMV from that of whole TMV by these physical chemical methods is very limited. Instead of trying to find the distribution by using any of these methods, a simple stripping model is proposed and the distribution is calculated theoretically. This calculation may give a reasonable estimate of the actual distribution.

In the stepwise removal of protein subunits by SDS, assume each protein subunit is stripped off with the same rate constant k. Then a series of steps can be written, each step removes one protein

subunit by k.

$$V_N \xrightarrow{k} V_{N-1} + P,$$

$$V_{N-1} \xrightarrow{k} V_{N-2} + P,$$

$$\cdots$$

$$V_{N-n+1} \xrightarrow{k} V_{N-n} + P, \qquad (5)$$

$$\cdots$$

$$V_2 \xrightarrow{k} 2P .$$

V_N designates the whole virus with N protein subunits. V_{N-n} is the partial virus with n protein subunits removed. N is the total number of protein subunits in the intact virus and has a value of 2,130. P is the protein subunit. Since the molar SDS concentration is so much higher than that of TMV, one assumes that every stripped protein subunit is completely surrounded by SDS and is unable to go back onto the virus. Therefore the back reactions are assumed to be negligible. For the same reason, these equations do not include SDS. (The dependence of the stripping rate on SDS concentration comes in the rate constant k only.) With this simple first order decay model, the distribution of the concentration of any virus with n protein subunits removed can readily be calculated as outlined in the following.

Rate equations are written one for each step in (5):

$$\frac{dV_N}{dt} = -k\ (V_N)\ ,$$

$$\frac{dV_{N-1}}{dt} = -k\,(V_{N-1}) + k\,(V_N),$$

$$\cdots$$

$$\frac{dV_{N-n}}{dt} = -k\,(V_{N-n}) + k\,(V_{N-n+1}), \qquad (6)$$

$$\cdots$$

$$\frac{dV_2}{dt} = -k\,(V_2),$$

where parentheses represent concentrations. The concentration of V_N at time t can be obtained by solving the first equation in (6):

$$(V_N) = V_N^o\, e^{-kt}$$

where V_N^o is the concentration of V_N at t = 0, i.e. the initial unstripped virus concentration. By substituting (V_N) into the second equation in (6), the concentration of V_{N-1} can be solved similarly as:

$$(V_{N-1}) = V_N^o\, kt\, e^{-kt}.$$

This calculation is carried successively to step n+1 in (6). Then the expression for the concentration of V_{N-n} is obtained:

$$(V_{N-n}) = \frac{V_N^o\, k^n t^n\, e^{-kt}}{n!}. \qquad (7)$$

Dividing (V_{N-n}) by $V_N{}^o$ gives the fraction of virus with n protein subunits removed (at time t) and equation (7) becomes

$$f_n = \frac{(V_{N-n})}{V_N{}^o} = \frac{k^n t^n\, e^{-kt}}{n!} \qquad (8)$$

Equation (8) was used to calculate the distribution of f_n for the three stripping experiments carried out in this work. For each extent of stripping, an effective time T = kt was chosen so that

$$\frac{\sum_n f_n n}{2,130} = \frac{\sum_n \frac{T^n e^{-T}}{n!} n}{2,130} = \text{extent of stripping.}$$

For each value of T so chosen, f_n values for all n (0-2,130) were calculated from

$$f_n = \frac{T^n e^{-T}}{n!} \qquad (9)$$

and the results were plotted vs. n as shown in Fig. 10. The half band width of each curve gives some ideas about how broad the stripping distribution is. As one can see the width becomes larger as the extent of stripping increases, which is also in agreement with the electron microscopic measurements. Based on the half band width, this simple model predicts a spread of at least 12, 19, and 23 protein subunits for the 1.19, 3.14, and 4.58% stripping respectively. Since there are three nucleotides for each protein subunit in TMV, this model actually predicts the widths of 36, 57, and 69 nucleotides for each case.

To improve the distribution in the lengths of the stripping products, it is best to start with homogeneous ^{32}P-TMV preparation. It has been shown by Symington et al. (1962) that short virus rods are natural products of TMV biosynthesis. Thus to remove those short rods each ^{32}P-TMV preparation was purified by passing through an agarose gel A150M column (II B) which sorted the viral rods according to their sizes. Only the homogeneous ^{32}P-TMV present in the center portion of the main peak was taken for subsequent stripping experiments. Not only should homogeneous ^{32}P-TMV preparation be used for stripping, the stripping conditions employed

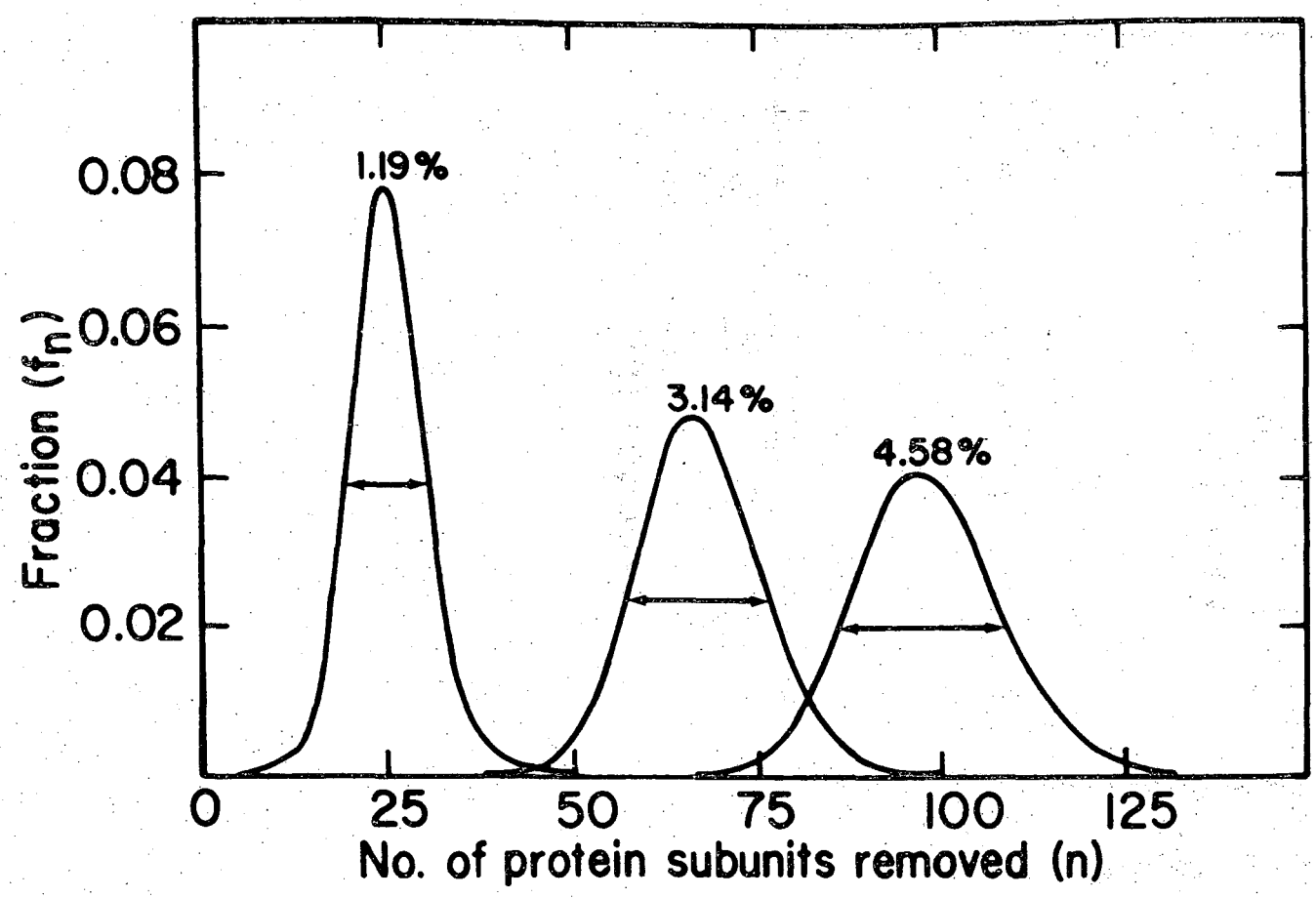


Fig. 10. Distribution of f_n as calculated from $f_n = \frac{T^n e^{-T}}{n!}$. T is 25.3, 66.9, and 97.5 for 1.19, 3.14, and 4.58% stripping respectively. Double arrow indicates half hand width.

should also be such as to maximize the homogeneity of the products. Neutral pH was used since either acid or base was shown to remove protein subunits of TMV at random positions (Fraenkel-Conrat, 1957; Schramm et al., 1955). Heat itself was shown to degrade TMV into its components (Cohen & Stanley, 1942) thus high temperature was avoided. All solutions were allowed to reach equilibrium at room temperature. Other chemical agents as urea (Stanley & Lauffer, 1939) and phenol etc. (Bawden & Pirie, 1940) should be undoubtedly excluded because of their ability to split TMV. The foregoing conditions required that stripping be strictly due to the SDS sequential removal of protein subunits from one end the viral rod. 0.002 M EDTA at neutral pH should always be present in the stripping mixture to prevent end to end aggregation of TMV (Steere, 1963). The presence of aggregated particles was believed to be the main cause of delaying the initiate stripping by SDS in some viral rods. Finally, the concentration of SDS and virus was adjusted so that the stripping rate was relatively fast yet completely controllable.

D. T_1-RNase Fingerprints of Exposed 3' Terminus of ^{32}P-TMV-RNA

The exposed 3' end of TMV-RNA obtained above was completely hydrolysed by T_1-RNase and the digestion products were fractionated by the two-dimensional ionophoresis procedure as described in II E and F. Figs. 11, 12 and 13 are the three T_1-RNase fingerprints obtained for the 1.19, 3.14 and 4.58% exposed 3' terminus of ^{32}P-TMV-RNA respectively. These will later be referred to as the partially stripped fingerprints. Fig. 14 is a map which shows the

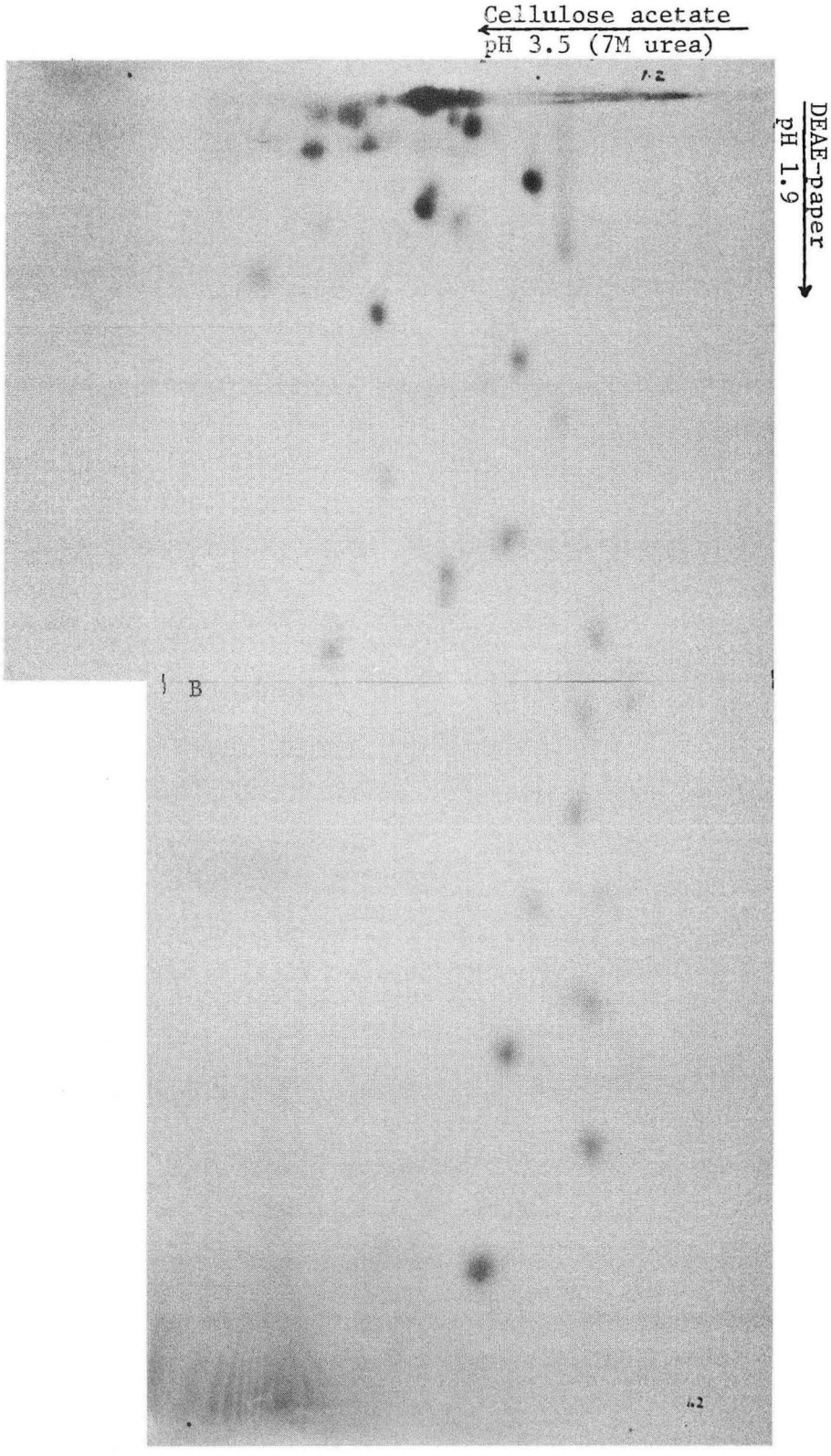


Fig. 11. Radioautograph of complete T_1-PNase digests of the 1.19% exposed 3' terminus of ^{32}P-TMV-RNA. B is the position of the blue marker on the second dimension.

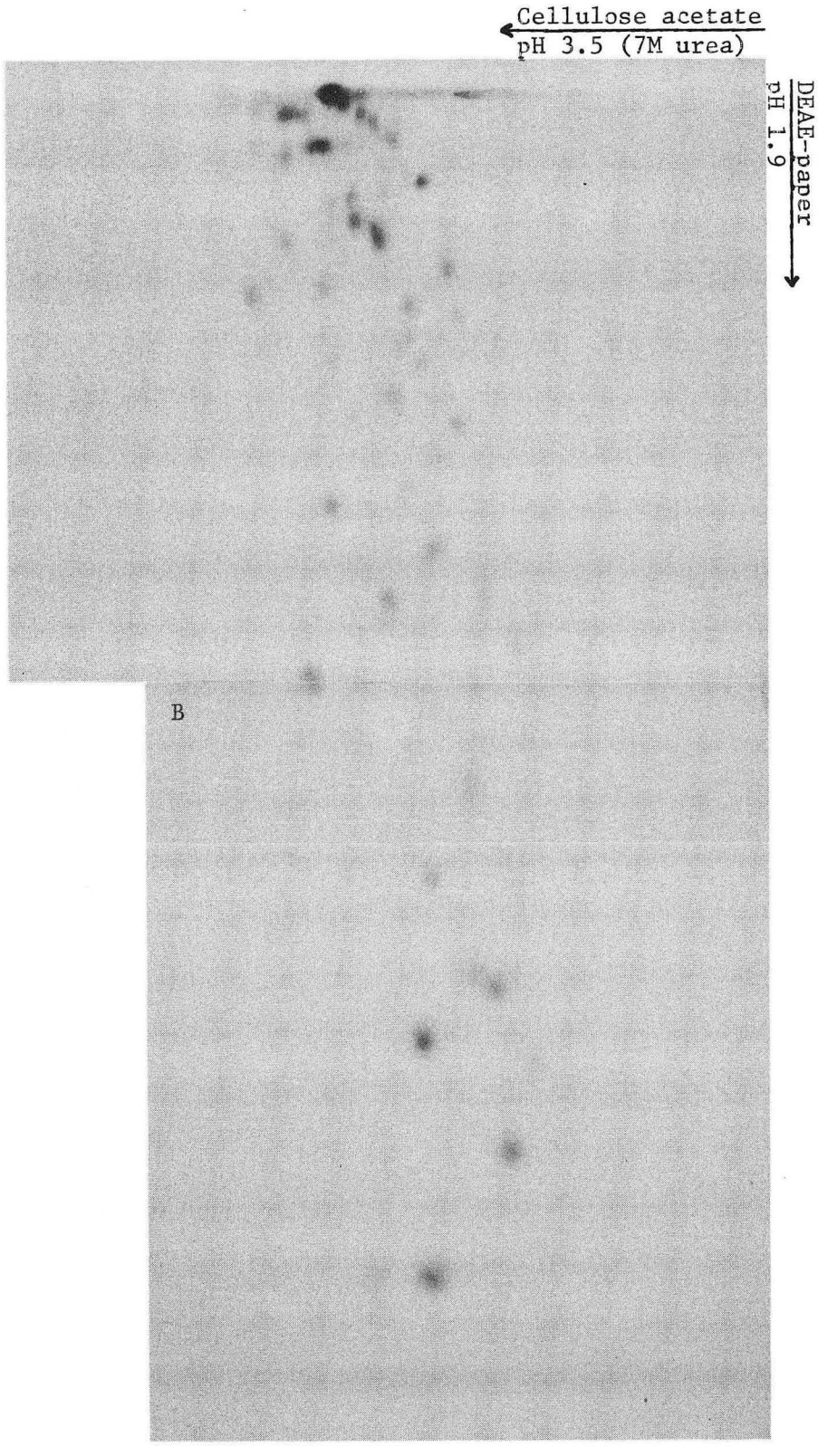


Fig. 12. Radioautograph of complete T_1-RNase digests of the 3.14% exposed 3' terminus of ^{32}P-TMV-RNA. B is the position of the blue marker on the second dimension.

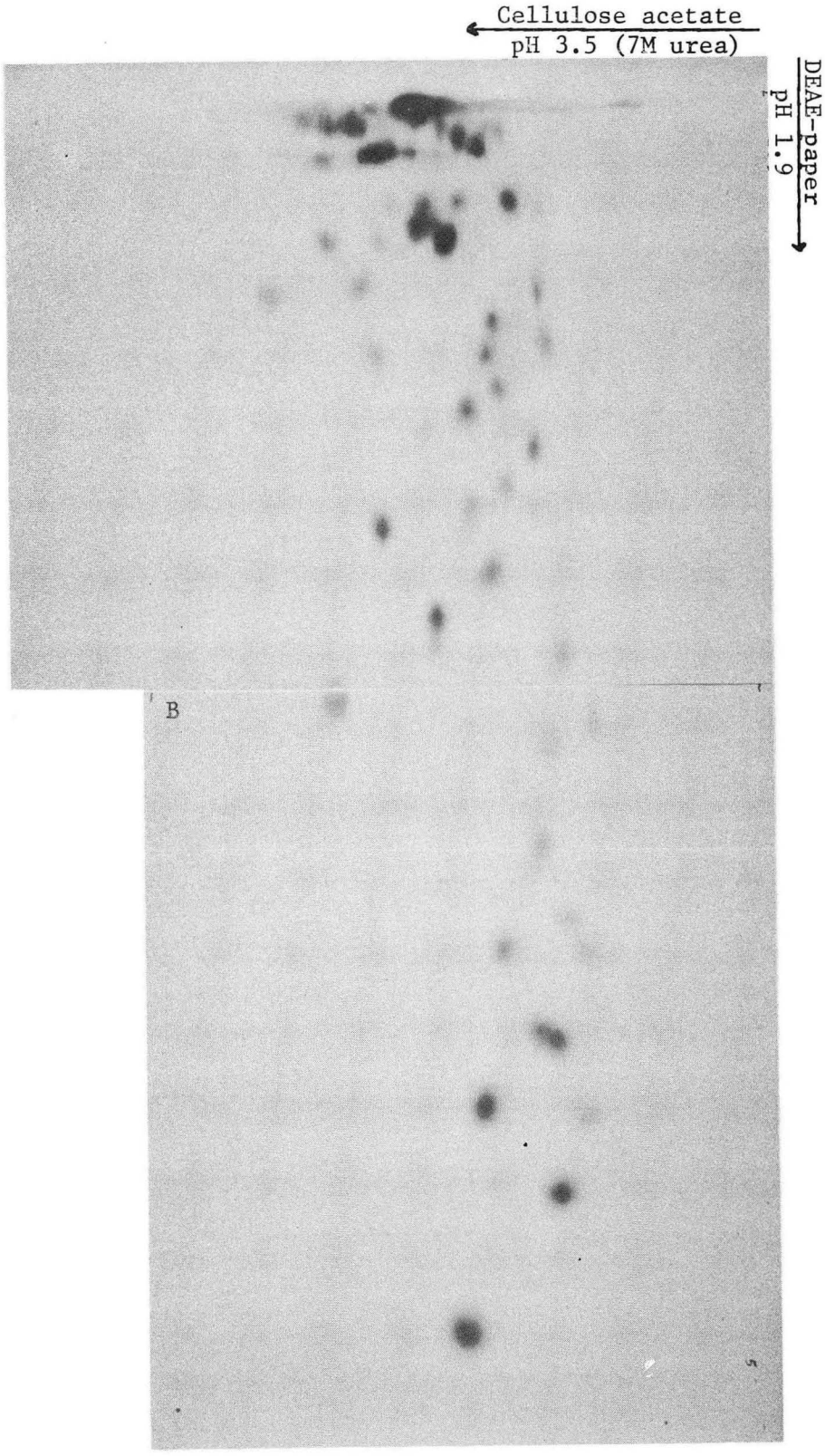


Fig. 13. Radioautograph of complete T_1-RNase digests of the 4.58% exposed 3' terminus of ^{32}P-TMV-RNA. B is the position of the blue marker on the second dimension.

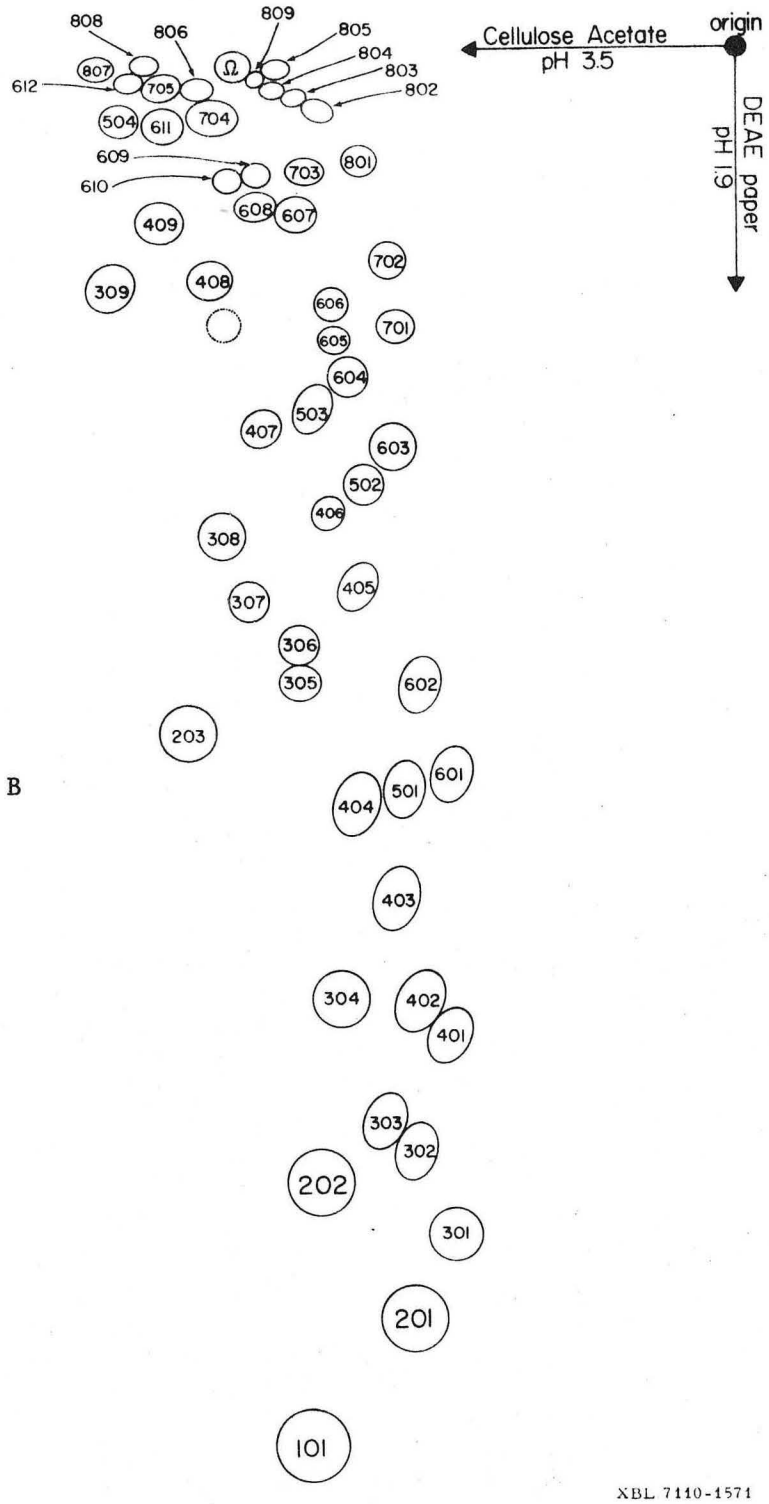


Fig. 14. The map showing the system of numbering of the spots in Figs. 11-13. B is the position of the blue marker on the second dimension.

system of numbering of the spots. For each extent of stripping two independent fractionations were carried out for its T_1-RNase digest. One was used to determine the amount of material present in each spot. The other was used for subsequent secondary enzyme hydrolysis.

(i) Structure of the oligonucleotides

The positions of the oligonucleotides from T_1-RNase digests fractionated on the DEAE-paper were determined from their radioautograph. Each spot was cut out and eluted with TEA-HCO_3^-. The oligonucleotides obtained were treated with PAN-RNase. The digestion products were fractionated by ionophoresis on DEAE-paper at pH 3.5. The experimental details were described in Materials and Methods (II G). The fractionation results are shown in Figs. 15-18. In each fractionation a sample of T_1-RNase digested ^{32}P-TMV-RNA, hydrolysed by PAN-RNase, was run along side as a control. The sequences of the separated oligonucleotides were thus determined according to the standard run described in Fig. 6. These sequences are also shown on the left of each radioautograph in Figs. 15-18.

The results obtained by PAN-RNase digestion were very clear. The digestion products were, in all cases, unambiguously separated and gave considerable information about the structure of the oligonucleotides from T_1-RNase digests. The amount of each product from PAN-RNase digests was estimated in the scintillation counter. Since the sequences of these products were known, the yield of a given mononucleotide can be expressed relative to one mole of G. Thus the base compositions of all the oligonucleotides from T_1-RNase

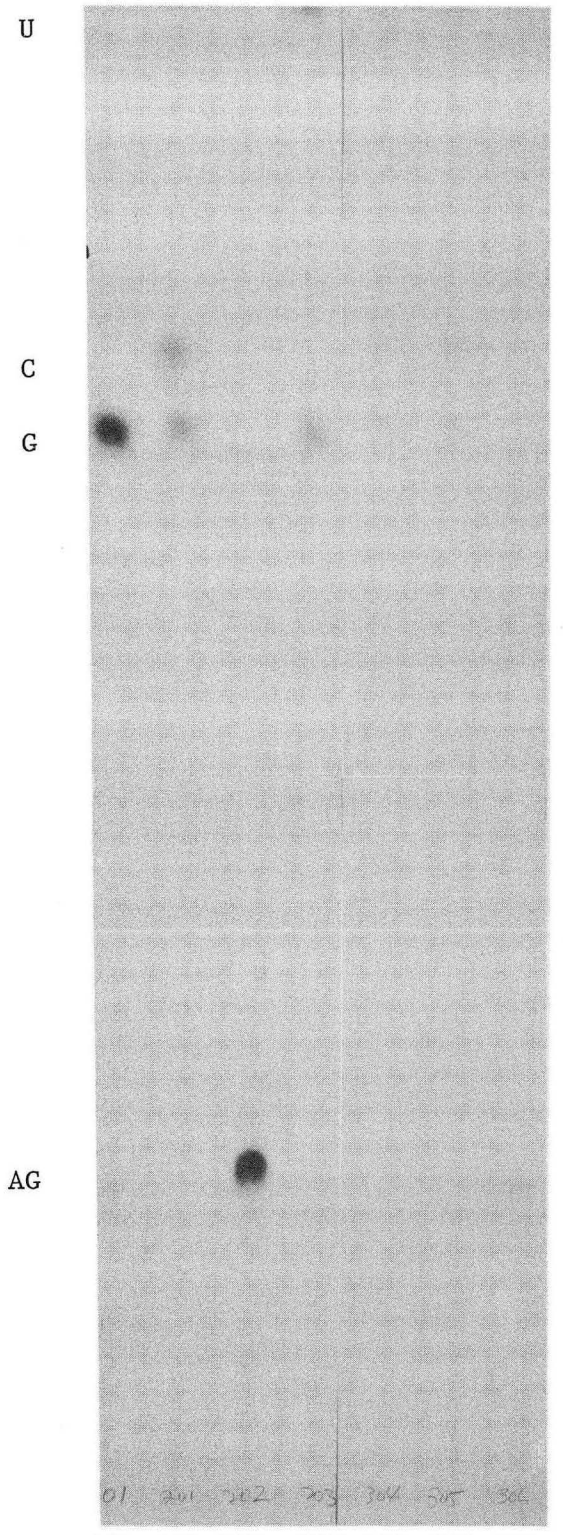


Fig. 15. Products from PAN-RNase digests of oligonucleotides from T_1-RNase digests. The numbers correspond to those in Fig. 14.

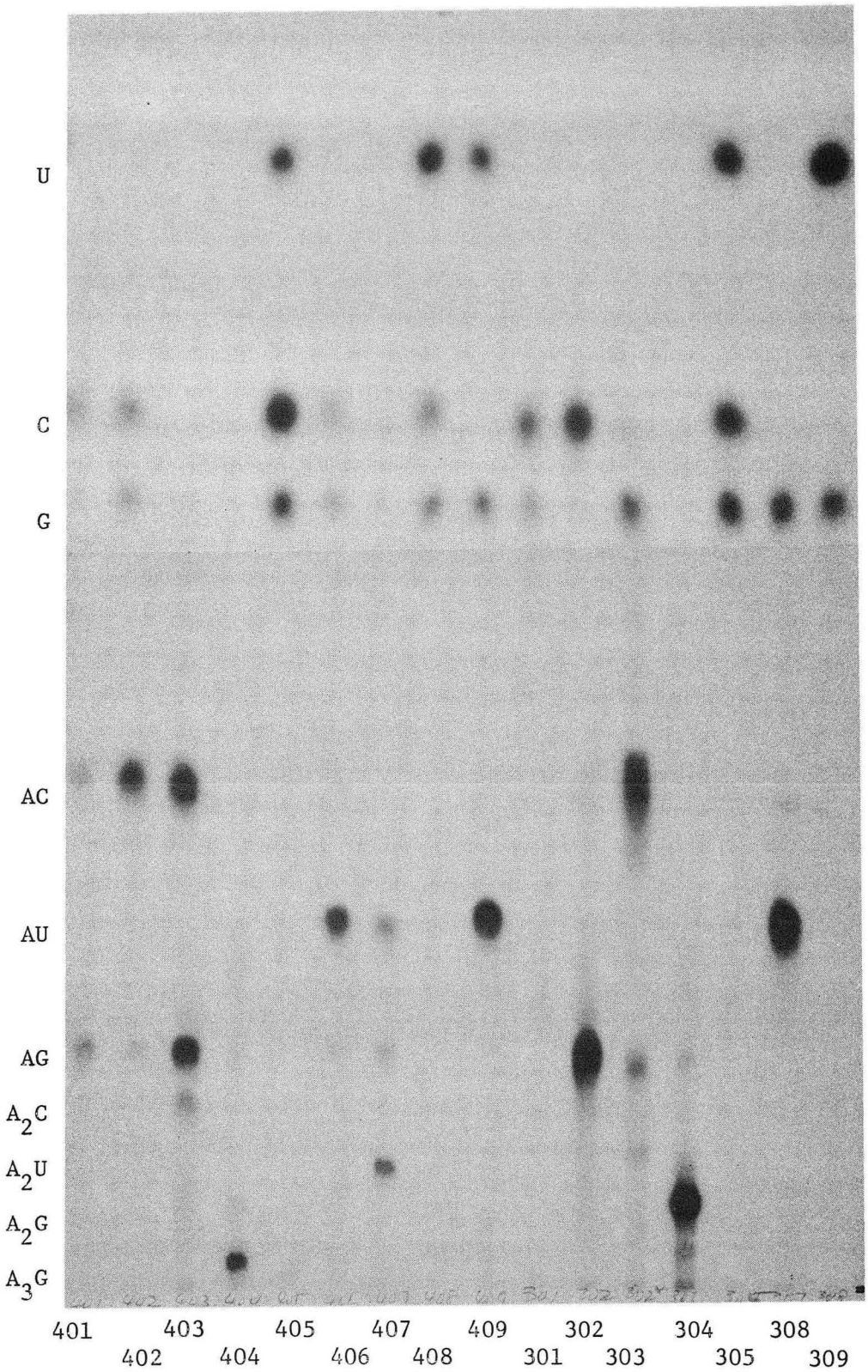


Fig. 16. Products from PAN-RNase digests of oligonucleotides from T_1-RNase digests. The numbers correspond to those in Fig. 14.

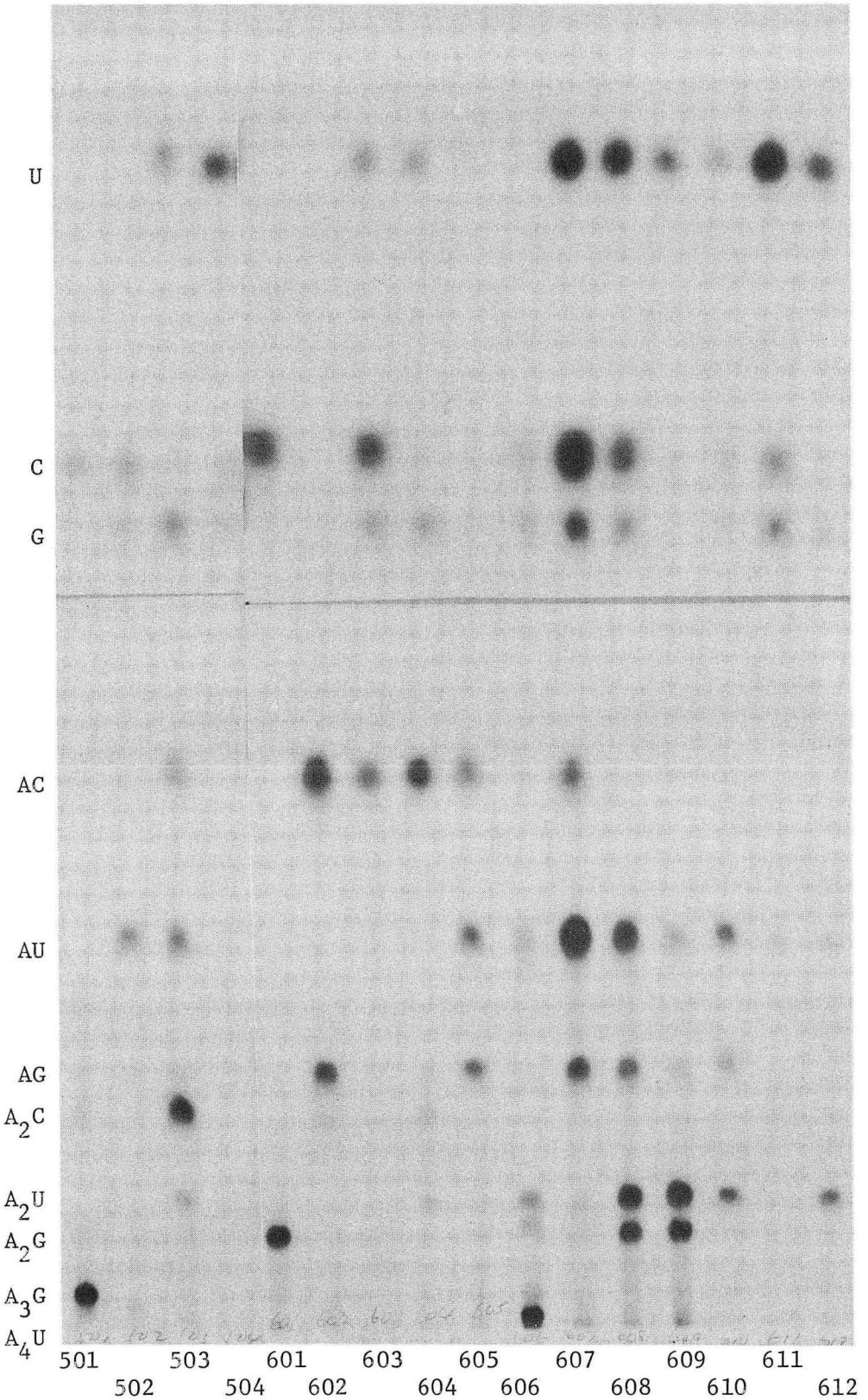


Fig. 17. Products from PAN-RNase digests of oligonucleotides from T_1-RNase digests. The numbers correspond to those in Fig. 14.

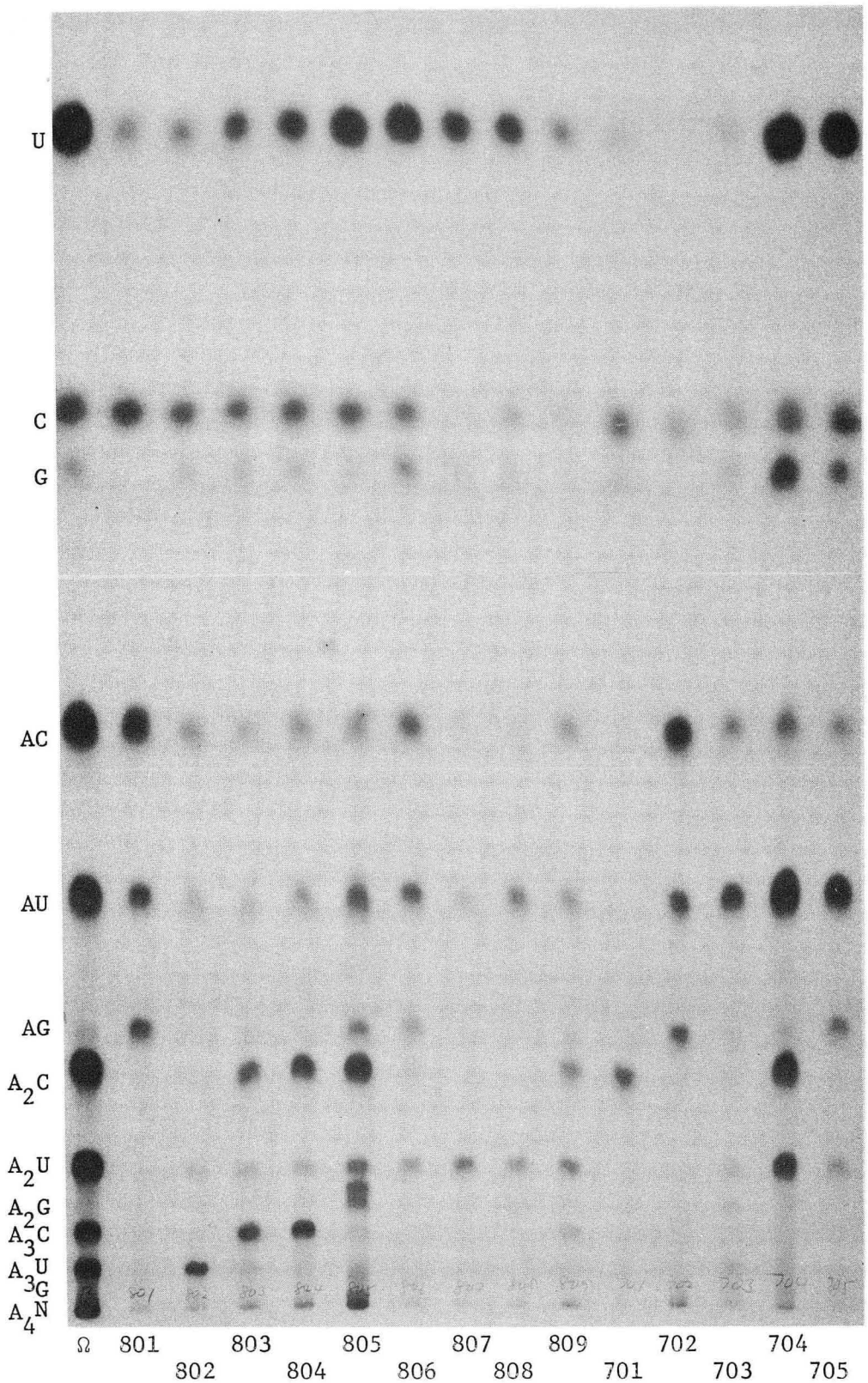


Fig. 18. Products from PAN-RNase digests of oligonucleotides from T_1-RNase digests. The numbers correspond to those in Fig. 14. N stands for C, U, and G.

0 0 0 0 3 8 0 0 3 1 1

digests were determined and the results are summarized in Table VII. An empirical formula as well as the chainlength for each oligonucleotide are also given in the table. The molar ratios of mononucleotides in most cases agreed with ± 10%, which indicated this is a good technique to determine the base composition. To prove the accuracy of these results, KOH hydrolysis was also carried out with some of the oligonucleotides. They were indicated with '*' in Table VII. The base compositions obtained from alkaline hydrolysis were the same as those obtained from PAN-RNase digests. Moreover, these were the compositions which would be reasonably predicted from their positions in the T_1-RNase fingerprint.

From the information of the oligonucleotide chainlength and the relative amount of the PAN-RNase digestion products, the structures of the oligonucleotides from T_1-RNase digests can be deduced. Each should terminate with a G at the 3' end. They are shown in the last column of Table VII. G(101) is the only mononucleotide from T_1-RNase digestion; this somewhat indicates good specificity of the nuclease. All di- and trinucleotides are well separated by the two-dimensional fractionation technique used in this work and their sequences are determined from PAN-RNase digestion. Spot 302 contains equal molar amounts of C and AG which gives the sequence CAG. Spot 303 contains G and AC as well as some C and AG. The relative amount of these nucleotides suggests the sequences of ACG and CAG. This indicates a slight overlapping of the two spots. Spot 304 contains almost exclusively AAG, unattacked by PAN-RNase. The small amount of AG can be explained as resulted from the secondary

Table VII

Oligonucleotides from T_1-RNase Digests of 3' Terminus of ^{32}P-TMV-RNA

Spot number (Fig. 14)	Base Composition C	A	G	U	Empirical formula	Chain-length n	PAN-RNase digestion products	Structure deduced
101	-	-	1	-	G	1	G	G
201	1.12	-	1	-	CG	2	C,G	CG
202	-	1	1	-	AG	2	AG	AG
203	-	-	1	1.01	UG	2	U,G	UG
301	2.11	-	1	-	C_2G	3	C,G	CCG
302	1.30	1	1	-	ACG	3	C,AG	CAG
303	0.89	0.89	1	-	ACG	3	AC,G	ACG
304	-	2	1	-	A_2G	3	AAG	AAG
305	0.98	-	1	0.91	UCG	3	U,C,G	UCG
306	1.19	-	1	1.12	UCG	3	U,C,G	CUG
307	-	1	1	1.06	AUG	3	AG,U	UAG
308	-	1.18	1	1.18	AUG	3	AU,G	AUG
309	-	-	1	2.35	U_2G	3	U,G	UUG
401	2.22	0.83	1	-	AC_2G	4	C,G,AC,AG	CCAG + (AC,C)G
402	2.32	1.13	1	-	AC_2G	4	C,G,AC	(AC,C)G
403	1.27	2.27	1	-	A_2CG	4	G,AC,AG,A_2C	ACAG + AACG
404	-	3	1	-	A_3G	4	AAAG	AAAG
405*	2.35	-	1	1.07	UC_2G	4	U,C,G	(U,C,C)G
406	1.17	1.00	1	1.00	ACUG	4	C,G,AU	(AU,C)G

Table VII (cont.)

Spot number (Fig. 14)	Base Composition C	A	G	U	Empirical formula	Chain-length n	PAN-RNase digestion products	Structure deduced
407	-	2.36	1	1.37	A_2UG	4	G,Au,AG, A_2U	AUAG + AAUG
408	1.11	-	1	2.26	CU_2G	4	U,C,G	(U,U,C)G
409	-	1.22	1	2.18	AU_2G	4	U,G,AU	(AU,U)G
501*	1.35	3	1	-	A_3CG	5	C,A_3G	CAAAG
502	2.00	0.87	1	0.87	AC_2UG	5	C,G,AU	(AU,C,C)G
503	0.94	1.89	1	1.21	A_2CUG	5	U,C,G,AC, AU,A_2C,A_2U	(A_2C,U)G,(A_2U, C)G,(AC,AU)G
504*	1.30	-	1	2.76	U_3CG	5	U,C,G	(U,U,U,C)G
601*	3.26	2.00	1	-	A_2C_3G	6	C,A_2G	CCCAAG
602*	2.21	3.20	1	-	A_3C_2G	6	AC,AG	ACACAG
603*	3.26	1.02	1	1.08	AC_3UG	6	U,C,G,AC	(AC,C,C, U)G
604*	1.80	1.80	1	0.96	A_2C_2UG	6	U,G,AC	(AC,AC,U)G
605*	1.12	3.39	1	1.27	A_3CUG	6	AC,AU,AG	(AC,AU)AG
606	1.20	3.86	1	1.14	A_4CUG	7	C,G,A_2U, A_2G,A_4U	(A_2U,C)A_2G (A_4U,C)G
607*	1.96	1.45	1	2.19	AC_2U_2G	6	U,C,G,AC, AU,AG	(C,C,U,U)AG (AC,C,U,U)G (AU,C,C,U)G
608*	1.15	1.94	1	2.18	A_2CU_2G	6	U,C,G,AU, AG,A_2U,A_2G	(C,U,U)AAG (AU,C,U)AG (AU,AU,C)G (AAU,C,U)G
609*	-	3.13	1	2.06	A_3U_2G	6	U,AU,AG, A_2U,A_2G	(AU,U)AAG (AAU,U)AG
610	-	2.16	1	2.32	A_2U_2G	5	U,G,AU, AG,A_2U	(AU,U)AG (AAU,U)G (AU,AU)G

Table VII (cont.)

Spot number (Fig. 14)	Base Composition C	A	G	U	Empirical formula	Chain-length n	PAN-RNase digestion products	Structure deduced
611	1.07	-	1	3.24	CU_3G	5	U,C,G	(U,U,U,C)G
612	0.79	2.07	1	5.10	A_2CU_5G	9	U,C,G,AU, A_2U	(AU,AU,U,U, U,C)G (AAU,U,U,U, U,C)G
701	5.19	1.82	1	1.12	A_2C_5UG	9	U,C,G,A_2C	(AAC,C,C,C, C,U)G
702	4.91	5.03	1	1.20	A_5C_5UG	12	C,AC,AU,AG	(AC,AC,AC, AU,C,C)AG
703	1.86	2.27	1	2.08	$A_2C_2U_2G$	7	U,C,G,AC, AU	(AU,AU,C,C)G (AU,AC,C,U)G
704	1.14	2.18	1	3.47	A_2CU_3G	7	U,C,G,AC, AU,A_2C,A_2U	(AAC,U,U,U)G (AAU,U,U,C)G (AC,AU,U,U)G (AU,AU,U,C)G
705	1.76	1.54	1	5.26	$A_2C_2U_5G$	10	U,C,G,AC, AU,AG,A_2U	
801*	4.19	3.32	1	2.03	$A_3C_4U_2G$	10	U,C,G,AC, AU,AG	
802	5.32	2.99	1	3.82	$A_3C_5U_4G$	13	U,C,G,AC, AU,A_3U	
803	4.22	3.73	1	3.77	$A_4C_4U_4G$	13	U,C,G,AC, AU,A_2C,A_3C	
804	5.97	5.10	1	5.25	$A_5C_6U_5G$	17	U,C,G,AC,AU, A_2C,A_3C	
805	2.85	3.92	1	5.02	$A_4C_3U_5G$	13	U,C,G,AC,AU, AG,A_2C,A_2U,A_2G	
806	1.97	1.56	1	4.70	$A_2C_2U_5G$	10	U,C,G,AC,AU,AG, A_2C,A_2U	
807	1.09	1.70	1	7.26	A_2CU_7G	11	U,C,G,AC,AU,AG, A_2U	

0 0 0 0 3 8 0 0 0 1 3

Table VII (cont.)

Spot number (Fig. 14)	Base Composition C	A	G	U	Empirical formula	Chain-length n	PAN-RNase digestion products	Structure deduced
808	1.78	2.22	1	6.22	$A_2C_2U_6G$	11	U,C,G,AC,AU, A_2U,A_2G	
809	5.62	7.92	1	6.95	$A_8C_6U_7G$	22	U,C,G,AC,AU, A_2C,A_2U,A_3C	
Ω	15.3	33.2	1	18.4	$A_{33}U_{18}C_{15}G$	67	U,C,G,AC,AU, A_2C,A_2U,A_3C	
by D. Lloyd	15	35	1	19	$A_{35}U_{19}C_{15}G$	70	U,C,G,AC,AU,A_2C, A_2U,A_3C	

*The base composition was also obtained from KOH hydrolysis.

split after A by PAN-RNase (Lloyd, Ph.D. thesis, 1969). Both spots 305 and 306 contain equal amount of U, C, and G. Although the exact sequence of each can be obtained by determining the 5' terminal residue, the sequences were assigned by comparison with those obtained by other workers (Fellner et al., 1970). The other four trinucleotides are completely resolved and their sequences are obtained without any difficulties. Of the nine spots of tetranucleotides obtained from the 3' terminus of ^{32}P-TMV-RNA, only spot 404 can be assigned as a single tetranucleotide AAAG. Spot 401 contains almost equal amount of CCAG and (AC,C)G; spot 402 contains mostly (AC,C)G. This indicates there is some overlapping between CCAG, ACCG and CACG. Spot 403 produces AAC, G, AC, AG whereas spot 407 produces AAU, G, AU, AG, when treated with PAN-RNase. The sequences deduced suggest the following pairs of sequence isomers are not separable: AACG and ACAG, AAUG and AUAG. For the other tetranucleotide spots all the probable sequences are shown, each spot may contain one or all of the isomers. There are six spots containing pentanucleotides. Only CAAAG(501) occurs as a single spot. The base compositions of the other pentanucleotide spots are determined, but the sequences have not been studied in detail. The probable sequences are shown and it seems likely that each of the spots is a mixture of isomers. There are two hexanucleotide spots, containing no U residue, each has a unique sequence. Spot 601 yields only C and AAG when treated with PAN-RNase. The molar ratio of C to AAG is 3, therefore the sequence must be CCCAAG. Spot 602 also contains only one hexanucleotide with the sequence ACACAG, since AC and AG are the only products from

PAN-RNase digests and their ratio is 2:1. Because of the increase of the number of products from PAN-RNase digests, it becomes more difficult to deduce the sequence of an oligonucleotide when its chain-length becomes larger. Only the probable structures are given for spots 603-704. For oligonucleotides of ten bases or longer only the base composition has been determined.

(ii) Molar yields of the oligonucleotides:

With the second two-dimensional fractionation, the positions of the oligonucleotides were determined from the radioautograph. Each spot was cut out, put in a vial containing 10 ml of scintillation fluid, and its ^{32}P activity (CPM_i) was counted in the scintillation counter. To determine the molarity of the oligonucleotides present in a spot one needs to know their chainlengths and the activity of one phosphate. The chainlength of the oligonucleotides in any spot has been determined from their base compositions. Due to the spread of the stripping distribution one is not sure whether a particular fragment of the RNA is totally exposed by SDS or not. Therefore one can no longer find the activity of a single phosphate by choosing a reference spot as done in III B. However, the sum over the activities of all spots is the total number of nucleotides exposed times the activity of one phosphate. The total number of nucleotides exposed regardless of the stripping distribution is equal to the average extent of stripping multiplied by the total number of nucleotides in the RNA molecule. The total chainlength was reported in this work for ^{32}P-TMV-RNA to be 6,284 ± 2.4%. Therefore the activity of one phosphate (Act(P)) was calculated as

$$Act(P) = \frac{\sum_i CPM_i}{\text{Average extent of stripping} \times 6{,}284}.$$

Then the molarity of the oligonucleotides present in any spot i can be obtained from

$$m_i = \frac{CPM_i}{Act(P) \cdot n}.$$

The resultant molar yields of these oligonucleotides from T_1-RNase digests are listed in Table VIII for each extent of stripping. The chainlength n is also listed. The values obtained from the whole ^{32}P-TMV-RNA (100% stripping) are listed in the last column. For spots which contain a single oligonucleotide, it is the total number of that oligonucleotide, e.g., AUG(308) can be found 52 times after a G residue in a TMV-RNA molecule. Otherwise, it is the total number of all oligonucleotides having the same base composition, e.g., tetranucleotides with the base composition A_2CG(403) occur 11 times after a G residue in a TMV-RNA molecule. Because of the difficulties in matching the longer oligonucleotides in the partially stripped fingerprints with that in the 100% fingerprint. Some molarities are not listed for 100% stripping.

The amount of oligonucleotides in any spot increases as more RNA is being exposed. If the amount of a given oligonucleotide is plotted against the percent stripping, ideally, a sigmoid curve should be obtained. The molar amount should reach a plateau at unity. For an oligonucleotide which only appears once, there should only be one plateau. The point at which 1/2 of this oligonucleotide has been observed is its location in the RNA. For an oligonucleotide which appears more than once, but widely separated in the RNA chain,

0 0 0 0 3 8 0 0 0 1 5

Table VIII

Molar Yields of Oligonucleotides present

in Spots of T_1-RNase Fingerprints

Spot number (Fig. 14)	Chain-length n	% Stripping			
		1.19%	3.14%	4.58%	100%
101	1	3.04	9.62	12.69	314
201	2	1.01	3.28	4.37	63
202	2	1.32	4.14	4.85	141
203	2	0.72	2.94	3.49	113
** 301	3	0.10	0.56	0.80	15
302	3	0.45	1.50	2.28	30
303	3	0.22	1.20	1.58	25
304	3	0.35	1.28	1.65	47
305	3	0.32	0.45	0.64	13
306	3	0.46	1.34	1.84	19
* 307	3	-	0.26	0.16	17
308	3	0.22	1.18	1.73	52
309	3	0.38	1.01	0.95	39
** 401	4	-	0.27	0.58	9
** 402	4	0.20	0.43	0.56	
403	4	0.43	0.79	1.00	11
* 404	4	-	0.19	0.35	7
405	4	0.54	0.83	1.11	13
** 406	4	-	0.32	0.77	9
* 407	4	0.063	0.19	0.19	12
** 408	4	0.067	0.45	0.69	23
** 409	4	0.10	0.54	0.76	11
501	5	0.27	0.34	0.56	6
* 502	5	-	0.20	0.39	2
** 503	5	0.11	0.63	0.93	19
* 504	5	-	0.23	0.12	18
* 601	6	0.14	0.35	0.45	4
* 602	6	0.23	0.35	0.47	3
** 603	6	0.13	0.42	0.59	
604	6	0.31	0.32	0.51	
** 605	6	-	0.23	0.51	
** 606	7	0.080	0.37	0.51	
607	6	0.20	1.36	2.08	
608	6	0.49	0.90	1.70	
609	6	0.21	0.50	0.78	
* 610	5	0.049	0.35	0.18	9
611	5	0.37	0.61	0.66	21
* 612	9	-	0.18	0.40	
* 701	9	-	0.16	0.31	
* 702	12	0.13	0.28	0.30	
** 703	7	-	0.23	0.52	

Table VIII (cont.)

Spot number (Fig. 14)	Chain-length n	% Stripping			
		1.19%	3.14%	4.58%	100%
704	7	0.24	1.58	2.39	
** 705	10	0.15	0.58	0.47	
801	10	0.36	0.42	0.62	
* 802	13	0.24	0.24	0.50	
* 803	13	0.071	0.23	0.45	
* 804	17	-	0.23	0.098	
805	13	0.23	0.71	0.62	
806	10	0.31	0.50	1.00	
* 807	11	-	0.13	0.13	
* 808	11	-	0.20	0.13	
* 809	22	-	0.21	0.22	
Ω	70	0.26	0.32	0.50	1

* and ** please refer to next section.

this type of sigmoid increasing should occur wherever this oligonucleotide is being exposed. The number of plateaus is equivalent to the number of times it exists in the exposed portion of the RNA. The point at which (n-1/2) moles of the oligonucleotide have been observed corresponds to the location of the nth occurrence during the course of stripping. It is desired to estimate the locations of all oligonucleotides this way and arrange them in the order of priority. However, due to the distribution in stripping and the possibility of the same oligonucleotides which occur very close to each other, no sigmoid curves are observed. Only in a few cases, e.g., spots 309 and 405, skew sigmoid curves are obtained. The majority show the behavior of continuous increasing in amount. The approximate location of an oligonucleotide can thus not be estimated from the point at which 1/2 mole of it has been observed. The following section describes a method which used the relative molar yields to locate the Ω-mer. It is hoped to use this method to locate all oligonucleotides.

E. Location of the Ω-mer

The uniqueness of the Ω-mer was first reported by Mandeles (1968). Its chainlength and PAN-RNase digestion products were studied by Lloyd and Mandeles (1970). In this work, substantial amount of the Ω-mer was found in all three different extents of strippings, indicating its location is extremely close to the 3' end of the ^{32}P-TMV-RNA. Because of its large size (70 bases, Lloyd, Ph.D. thesis) and lack of G residues, it is of interest to determine its exact location. The following describes a method for this purpose:

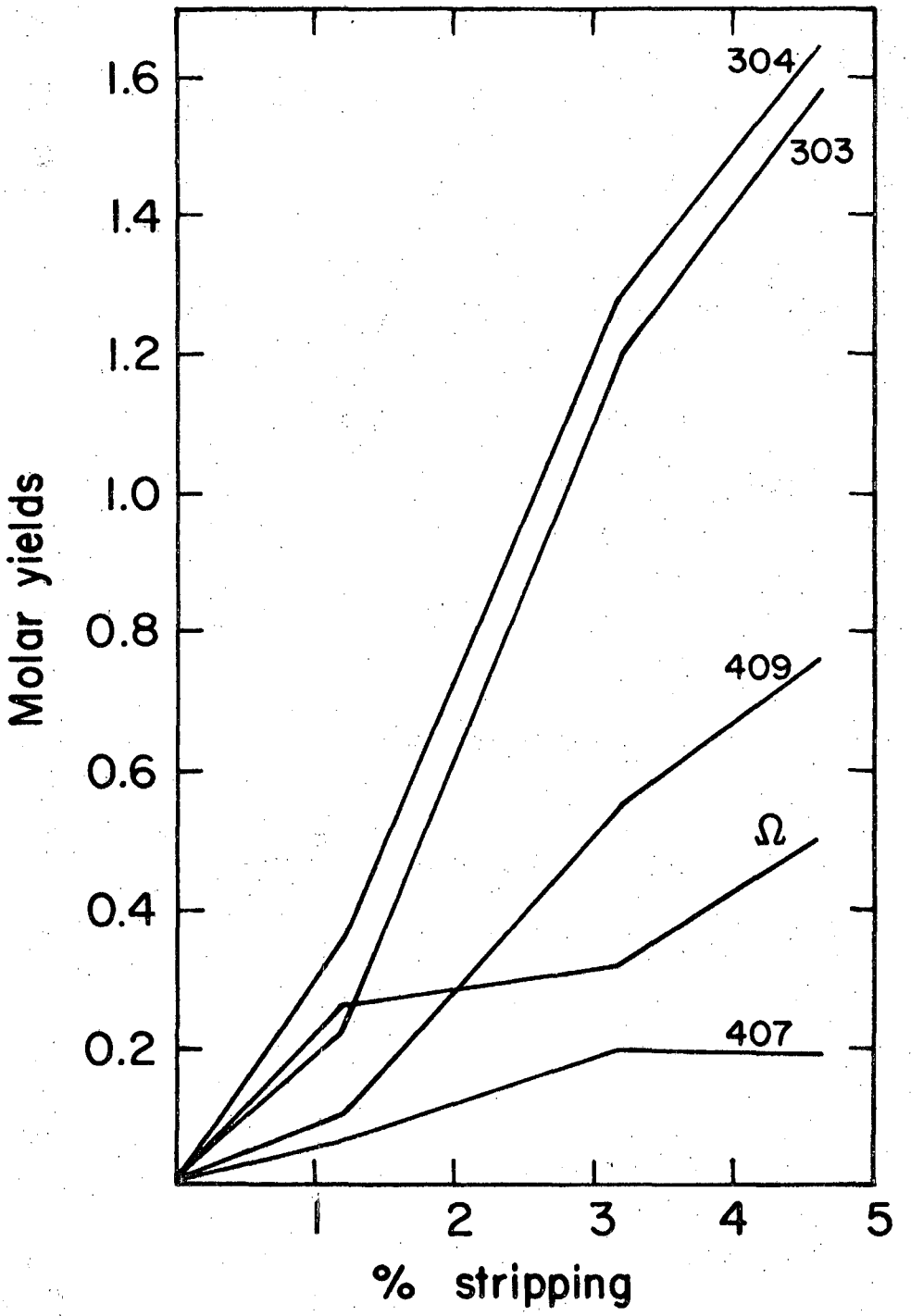
(a) find all possible oligonucleotides that occur prior to the Ω-mer.

(b) determine the number of times each such oligonucleotide occurs prior to the Ω-mer.

Any oligonucleotide that is found closer towards the 3' end than the Ω-mer is said to occur prior to the Ω-mer. Others are said to occur behind the Ω-mer.

When protein subunits of TMV are stripped sequentially from the 3' end by SDS, those oligonucleotides that occur prior to the Ω-mer are exposed ahead of the Ω-mer. When treated with T_1-RNase their molar yields must be higher than that of the Ω-mer. On the other hand, those oligonucleotides that occur behind the Ω-mer are not exposed unless the Ω-mer has already been exposed. Their molar yields must be lower than that of the Ω-mer. This fact is complicated by the oligonucleotides that occur multiple times in the RNA.

By plotting the molar yields in Table VIII vs. the percent stripping, a series of curves, showing how the amount of the oligonucleotides in each spot increases with the extent of stripping, are obtained. Fig. 19 shows some of these curves. Taking the curve of the Ω-mer as a reference, the curves can be grouped into three categories. The first is a group of curves below the reference curve (e.g. spot 407). Because of the lower molar yields, these curves correspond to the oligonucleotides that occur only once behind the Ω-mer. (If there is a possibility that some of these might occur twice or more behind the Ω-mer, they have to be far behind so that very little of each is exposed.) They are



XBL 722-278

Fig. 19. Diagram showing the relation between molar yields of the oligonucleotides from T_1-RNase digests and the percent stripping. Numbers are those in Table VIII.

represented by '*' in Table VIII. The second group, represented by '**' in Table VIII, consists of curves that intersect the reference curve (e.g. spot 409). All oligonucleotides in this category are either not exposed at all or less than 15% exposed at 1.19% stripping. As compared to the 26.2% exposure of the Ω-mer, the possibility of any of these oligonucleotides occurring prior to the Ω-mer is ruled out. The increase in the molar amount which overtakes that of the Ω-mer, consequently causing the curves to intersect, is due to the exposure of more than one such oligonucleotide behind the Ω-mer. The multiple exposure of the oligonucleotide behind the Ω-mer may well be its unresolved sequence isomers. Furthermore, the molar yields of these spots at 4.58% stripping are in all cases less than 95%. At this extent of stripping exposure of two sequence isomers behind the Ω-mer should be able to account for its molar yield. Spots 303, 308, 607, 609, 704, and 805 are exceptions in this category for two reasons: (1) their molarities at 1.19% stripping are close to that of the Ω-mer, (2) the molar amount becomes so large at greater extent of stripping, and would require exposure of a large number of these oligonucleotides behind the Ω-mer. Therefore, they probably should be classified in the next group. The third group consists of curves which occur above the reference curve (e.g. spot 304). To have molarity greater than that of the Ω-mer throughout the whole spectrum of stripping, at least one oligonucleotide from each of these spots should occur prior to the Ω-mer. The rapid increase in the molarity of an oligonucleotide with the extent of stripping suggests that this oligo-

0 0 0 0 3 8 0 0 0 1 8

nucleotide may occur more than once or its sequence isomer is present as well.

To determine the number of times an oligonucleotide may occur prior to the Ω-mer, the molar ratios of these oligonucleotides to the Ω-mer at each extent of stripping are calculated and listed in Table IX. A molar ratio of less than 2 at any extent of stripping probably indicates that only one oligonucleotide from that spot occurs once prior to the Ω-mer. These spots are listed as '1' in column 6 of Table IX. Those having all three molar ratios about one (305, 501, 604, and 801) are the oligonucleotides which are located closer to the Ω-mer, because their exposure brings immediate uncovering of the Ω-mer. On the other hand, oligonucleotides very near the 3' end are exposed at a small extent of stripping. As stripping progresses very little more are exposed, while the amount of the Ω-mer exposed increases. Consequently, this brings about an increasing molar ratio followed by a drop (304, 309, 611, and 805). The oligonucleotides whose molar ratios increase and reach a value about two (609 and 806) are the intermediate ones. The high molar ratios at large extent of stripping of some oligonucleotides (303, 308, 607, and 704) are due to the uncovering of these oligonucleotides, or their sequence isomers, behind the Ω-mer. In order for a given oligonucleotide to be present a second time, its molar ratio to the Ω-mer must be greater than 2.0 at every extent of stripping. Five spots are chosen and listed as '2' in column 6 of Table IX. Because their molar ratios remain unchanged at about two for every extent of stripping, spots 403 and 405, each containing two oligo-

Table IX

Molar Ratio of the Oligonucleotide to Ω-mer

Spot number (Table XIII)	Chain-length n	1.19%	3.14%	4.58%	No. of occurrence
101	1	11.6	30.25	25.55	11-12
201	2	3.86	10.30	8.79	3-4
202	2	5.03	13.02	9.76	5
203	2	2.75	9.25	7.01	2-3
302	3	1.72	4.72	4.60	2
303	3	0.86	3.77	3.18	1
304	3	1.35	4.02	3.31	1
305	3	1.23	1.41	1.30	1
306	3	1.74	4.22	3.70	2
308	3	0.82	3.70	3.48	1
309	3	1.43	3.15	1.90	1
403	4	1.65	2.50	2.01	2
405	4	2.06	2.60	2.24	2
501	5	1.02	1.06	1.12	1
604	6	1.18	1.01	1.03	1
607	6	0.77	4.26	4.19	1
608	6	1.87	2.82	3.43	2
609	6	0.82	1.56	1.57	1
611	5	1.42	1.92	1.33	1
704	7	0.93	4.96	4.81	1
801	10	1.37	1.31	1.26	1
805	13	0.86	2.22	1.25	1
806	10	1.18	1.57	2.01	1
				Total nucleotides	154-159

nucleotides (identical or sequence isomers), should be located quite close to the Ω-mer. The large molar ratios at 3.14% and 4.58% stripping for spots 302 and 306 indicate that at least one from each spot is located near the end and the others somewhere between the intermediate and the 3' terminal position. The continuous increase in molar ratio of spot 608 suggests the following possibilities of two locations: (1) both of them in the intermediate position, or (2) one close to the Ω-mer while the other close to the 3' end. Finally, to determine the number of times the mononucleotide G(101) and dinucleotides CG(201), AG(202), UG(203) occur prior to the Ω-mer, the lowest molar ratio among the three is chosen. It is impossible, however, to locate their position more closely.

From the chainlengths and the number of times all of these oligonucleotides occur prior to the Ω-mer, the location of the Ω-mer can be obtained. It should be mentioned here that a particular tetranucleotide should also be included. Since the 3' terminal nucleotide sequence -GCCCA has been determined independently by Steinschneider & Fraenkel-Conrat (1966) and Mandeles (1967). Upon T_1-RNase hydrolysis, tetranucleotide CCCA should be obtained. Due to the lack of terminal phosphate and residues U and G, it has not been located in the fingerprints obtained in this work. However, it must be included in determining the location of the Ω-mer. Therefore the 3' terminal residue G of the Ω-mer is 158-163 nucleotides from the 3' end of the TMV-RNA, and all of those T_1-RNase digestion products prior to the Ω-mer are:

11-12 G, 3-4 CG, 5 AG, 2-3 UG, 2 CAG, 1 ACG, 1 AAG, 1 UCG, 2 CUG, 1 AUG, 1 UUG, 2 $(A_2,C)G$, 2 $(U,C_2)G$, 1 CAAAG, 1 $(A_2,C_2,U)G$, 1 $(A,C_2,U_2)G$, 2 $(A_2,U_2,C)G$, 1 $(A_3,U_2)G$, 1 $(C,U_3)G$, 1 $(A_2,C,U_3)G$, 1 $(A_3,C_4,U_2)G$, 1 $(A_4,C_3,U_5)G$, 1 $(A_2,C_2,U_5)G$, and 1 CCCA.

The location of the Ω-mer determined above is by all means the upper limit. For a given oligonucleotide which occurs n times prior to the Ω-mer, it is completely detectable because its molar yields at any extent of stripping should be at least n times that of the Ω-mer. Even if this oligonucleotide is present in a spot, which is a mixture of chainlength and sequence isomers, it can be detected. Because, in this case, the total molar yields of all oligonucleotides in this spot would be much larger than that of the Ω-mer. It can then be concluded that no other oligonucleotide, besides the above listed, occurs prior to the Ω-mer. The uncertainty in the location of the Ω-mer may be due to oligonucleotides that produce unresolved spots. A spot of longer oligonucleotides is likely to contain a large number of unresolvable sequence isomers, all of which may actually occur behind the Ω-mer. However, the total molar yields of these sequence isomers may be larger than that of the Ω-mer. If this was found true at any extent of stripping carried out in this work, at least one of these isomers is placed prior to the Ω-mer. The error introduced this way could only locate the Ω-mer further away from the 3' end of RNA. Therefore, it can be concluded that the 3' terminal residue G of the Ω-mer is no more than 158 to 163 nucleotides from the 3' end of the TMV-RNA.

This result is consistent with that obtained from the approximate location of these oligonucleotides. Their molar yields are plotted vs. the percent stripping as shown in Fig. 19. Each curve is extrapolated to the percentage of stripping at which 1/2 mole of the oligonucleotide is exposed. The corresponding percent strippings (P(1/2)), which would estimate the approximate positions of these oligonucleotides, are summarized in Table X. The value for the Ω-mer is also listed and called M(Ω) for convenience. In general, any oligonucleotides with P(1/2) values smaller than M(Ω) occur prior to the Ω-mer. Those with P(1/2) comparable to M(Ω) are the oligonucleotides occurring close to the Ω-mer. On the other hand, those with P(1/2) much smaller than M(Ω) are the oligonucleotides close to the 3' end of the RNA. Thus the data in Table X indicates all the listed oligonucleotides occur prior to the Ω-mer. For spots 302, 306, 403, 405, and 608, the percent strippings at which 1.5 moles of the oligonucleotides are exposed are also given in Table X. These percent strippings estimate the approximate positions at which each of these oligonucleotides is exposed for the second time. Since they are smaller than M(Ω), the second occurrence is then also prior to the Ω-mer. Furthermore, the relative positions of these oligonucleotides with respect to the Ω-mer agree with those obtained previously. Those closer to the Ω-mer are oligonucleotides from spots 305, 403, 405, 501, 604, 608, and 801. Those closer to the 3' end are from spots 302, 304, 306, 309, 403, 405, 608, 611 and 805. Those in between are from spots 609, 806, 302, and 306. Oligonucleotides from spots 303, 308, 607, and 704 are predicted to be

Table X

The Approximate Location of the Oligonucleotides

from T_1-RNase Digestion

Spot number	Position at which 1/2 mole is exposed (in % stripping)	Position at which 1 1/2 mole is exposed (in % stripping)
101	0.19	
201	0.60	
202	0.43	
203	0.79	
302	1.28	3.17
303	1.73	
304	1.50	
305	3.53	
306	1.29	3.59
308	1.77	
309	1.59	
403	1.56	4.77
405	0.88	4.04
501	4.22	
604	4.48	
607	1.69	
608	1.26	4.21
609	3.15	
611	2.24	
704	1.56	
801	3.70	
805	2.31	
806	3.14	
Ω	4.61	

close to the 3' end. The only discrepancy is found in spots 403 and 405.

IV. DISCUSSION

Determination of nucleotide sequence in a nucleic acid involves methods for its chemical or enzymatic degradation, reliable techniques for the separation and structural analysis of the fragments produced, and for the determination of the order in which these fragments occur in the nucleic acid. Various methods have been well summarized by RajBhandary & Stuart (1966). A distinct type of sequential analysis is chemical or enzymatic stepwise degradation. The stepwise degradation produces an identifiable base, nucleoside or nucleotide from one end of the molecule. Steinschneider & Fraenkel Conrat (1966) have determined the sequence of five bases at the 3' end of TMV-RNA by using a chemical stepwise degradation method, which involves the oxidation of the terminal nucleotide by periodate. The main limitation in this method is that each degradation step is not completely quantitative. Thus application of this stepwise method is best only for small oligonucleotides. To degrade a nucleic acid into specific fragments requires highly specific enzymes. The best characterized ribonucleases are T_1-RNase, which is specific for cleavage of a phosphodiester bond after G, and PAN-RNase, which cleaves phosphodiester bonds after nucleotides U or C. Both nucleases produce oligonucleotides bearing 3' phosphates. A two-dimensional high voltage ionophoresis technique described by Sanger *et al*. (1965) has facilitated the separation and identification of these fragments obtained from either T_1- or PAN-RNase digests. Hence the main difficulty in determining the nucleotide

sequence lies in the ordering of these fragments. One commonly used method for arranging these fragments is the overlap method. This method determines the nucleotide sequence by comparing the overlapping regions of nucleotide fragments obtained from at least two different enzymatic degradations of the RNA. In order to be able to deduce a unique sequence for the entire RNA, enough overlaps between these fragments are required. Therefore overlap methods are usually carried out with fragments from partial degradation by ribonucleases. The partial degradation yields many similar, but not identical, fragments. The limitation of the overlap method is then in the difficulties of isolating large fragments, accurately estimating their amounts, and determining their complete nucleotide sequences. In regard to the limitations of the above two mentioned methods, Mandeles & Tinoco, Jr. (1963) have proposed a method which involves the formation of derivatives specific for either the 3' or 5' end of a nucleic acid. By controlled partial enzymatic hydrolysis, each molecule is broken into only two fragments. The halves carrying the end label are isolated and separated according to chainlength. By studying the nucleotide sequence of the unlabeled end of each half fragment and counting the distance of it from the label permits one to obtain a complete sequence of a nucleic acid. This method does not require any overlap information nor any stepwise degradation procedure. Yet application to the determination of nucleotide sequence in molecules of the size of TMV-RNA is limited by the need to separate very long fragments. Even for just 1 to 5% of the

3' terminal end of TMV-RNA, one must distinguish between oligonucleotides of the order of a few hundred bases.

A method is introduced in this work for the determination of nucleotide sequences in the 3' end of TMV-RNA. It is made feasible by the characteristic SDS polar stripping of TMV. SDS progressively exposes the 3' end of TMV-RNA. The RNA exposed to different extents is hydrolysed by T_1-RNase and its products are analysed by the two-dimensional ionophoresis fractionation technique. The ordering of the oligonucleotides is facilitated by observing the extent of SDS stripping on their appearance. For an ideal SDS stripping distribution, the appearance of a given oligonucleotide is all or none, which gives its exact location in the RNA. Due to the wide spread of the stripping distribution, however, the ordering of the oligonucleotides cannot be simply obtained from their appearance. But the ordering can be obtained from a series of studies on their relative locations. The first step is to choose a particular oligonucleotide as a reference. Any oligonucleotide that occurs closer to the 3' end of the RNA than the reference is uncovered by SDS first. Hence its recovery in molar amount must be greater than that of the reference at any extent of stripping. With this criterion, the relative location of the oligonucleotides with respect to the reference can be obtained from a plot of their molar yields vs. the percent stripping. In this work the Ω-mer is chosen as a reference because of its uniqueness; it only occurs once along the whole RNA chain and it is the largest fragment obtained from complete T_1-RNase digestion of TMV-RNA (Mandeles,

1968). All oligonucleotides are located in relation to the Ω-mer. From a plot of the molar yields vs. the percent stripping, a given oligonucleotide is placed either prior (closer to the 3' end of the RNA) or behind (farther away from the 3' end of the RNA) the Ω-mer by comparing its rate of appearance to that of the Ω-mer. As a result, all oligonucleotides are separated into two groups by the Ω-mer. By counting the total number of oligonucleotides that occur prior to the Ω-mer, the location of the Ω-mer is obtained. It has been shown in this work that 45 to 48 T_1-RNase digestion products occur prior to the Ω-mer. This places the 3' terminal G of the Ω-mer 158 to 163 bases from the 3' end of the TMV-RNA.

The next step is to choose a second reference oligonucleotide and carry out the study of the relative location of the oligonucleotides as described for the Ω-mer. By finding all those oligonucleotides that occur prior to this reference, the location of this reference oligonucleotide can then be obtained. If this type of relative location study can be carried out for all the T_1-RNase digestion products that occur prior to the Ω-mer, the ordering of these oligonucleotides becomes possible.

The multiple occurrence of the oligonucleotides from T_1-RNase digestion is the main limitation of this method. Since the frequency of occurrence is greater for smaller oligonucleotides, the difficulty arises especially when a small oligonucleotide is chosen as a reference. The immediate approach to this problem is to control the SDS stripping so that it exposes very small amount of the RNA at a time. In this work, three different extents

of stripping have been obtained. They extend over a range of more than three hundred bases. The difference between two different extents of stripping corresponds to over a hundred bases. To order the oligonucleotides more precisely smaller interval of the extent of stripping is required. An ideal case is to uncover 20 to 30 bases (less than 0.5% stripping) at each extent of stripping. This corresponds to an increase of only a small number of oligonucleotides when hydrolysed with T_1-RNase. It has been shown in this work that the rate of SDS stripping decreases as the SDS concentration decreases. On the other hand, the rate is enhanced by heat (Symington & Commoner, 1967). Although the kinetics of the SDS stripping of TMV is not known, it seems reasonable that the stripping can be controlled by employing SDS concentration lower than 0.5% and temperature near 0° C, in a rate which can be stopped at short intervals. This enables one to deal with only a small segment of the terminal RNA within which the oligonucleotide, chosen as a reference, occurs only once. Once the location of the first occurrence is known for a given oligonucleotide, the difficulty of locating its second occurrence is then greatly diminished. Furthermore, hydrolysis of the exposed RNA by a second ribonuclease will aid the ordering of small oligonucleotide. PAN-RNase digestion may help greatly in ordering mononucleotide G and the three dinucleotides from T_1-RNase digestion.

The structural analysis of the oligonucleotides present in the spots of the partially stripped fingerprints (Fig. 11 to 13) has yielded information on their base compositions and PAN-RNase digestion products. The results on the base composition fit very

well with their positions in the fingerprints. As has been suggested by Sanger et al. (1965), lines can be drawn connecting oligonucleotides from T_1-RNase digestion with increasing number of residue A or C. From the base composition of the oligonucleotides obtained in this work, lines can be drawn in that manner. In fact, all oligonucleotides containing two U residues or less can be fitted in one of these lines. Examples of the lines that connect oligonucleotides with increasing number of residue A are those passing (Fig. 14): spots 101(G), 202(AG), 304(A_2G), and 404(A_3G); spots 305, 306(CUG), 406(ACUG), 503(A_2CUG), 605(A_3CUG), and 606(A_4CUG); spots 309(U_2G), 409(AU_2G), 610(A_2U_2G), and 609(A_3U_2G). Examples of the lines that connect oligonucleotides with increasing number of residue C are those passing (Fig. 14): spots 202(AG), 302, 303(CAG), and 401, 402(C_2AG); spots 307, 308(AUG), 406(CAUG), 502(C_2AUG), and 603(C_3AUG); spots 407(A_2UG), 503(CA_2UG), 604(C_2A_2UG), and 701(C_5A_2UG). There is no crossing of lines within each category. Furthermore, these results are in good agreement with those reported for the corresponding T_1-RNase fragments obtained from 5S rRNA (Brownlee & Sanger, 1967), 16S rRNA (Fellner et al., 1970), R17 RNA (Adams et al., 1969; Steitz, 1969). The results obtained from PAN-RNase digestion have shown that the fractionation procedure, which used high voltage ionophoresis in both dimensions, is not capable of completely separating sequence isomers. Only dinucleotides and trinucleotides are completely separated. Sequence isomers of oligonucleotides of 4 bases or longer usually occur in one large spot.

Some separation is found in spots 401 and 402, both contain oligonucleotides of base composition AC_2G. Spots 504 and 611 also appear to be sequence isomers since both contain oligonucleotides with base composition U_3CG. The same is found true for spots 705 and 806, where base composition $A_2C_2U_5G$ is found for both spots. Since the number of sequence isomers increases so rapidly as the oligonucleotide becomes larger, even though two spots are found for the same base composition, each of these spots may also be a mixture of sequence isomers. To further study the base sequence of oligonucleotides present in a spot, experiments should be carried out to determine the 5' terminal residue. This residue can be identified by treatment with alkaline phosphomonoesterase followed by snake venom phosphodiesterase. The former removes the terminal 3' phosphate. The latter breaks the 3' phosphoester bond so that the 5' terminal residue is present as a nucleoside. Since ^{32}P is used as a label it can then be separated and identified as the missing residue provided the base composition is known. A knowledge of the 5' terminal residue can eliminate a large number of sequence isomers that might be present in a spot. For sufficiently simple oligonucleotides, a unique sequence may well be deduced. In most cases, informations from partial degradation by snake venom phosphodiesterase or spleen phosphodiesterase are required to deduce a unique sequence. The details of experimental procedure will not be included here since they have been discussed by Sanger _et al_. (1965), and Brownlee & Sanger (1967).

0 0 0 0 3 8 0 0 3 2 2

The last point that needs mentioning is the total chain-length of TMV-RNA reported in this work. The chainlength of 6,284 ± 2.4% has been determined for the whole TMV-RNA. This result supports the data obtained by other chemical method (Caspar, 1963).

REFERENCES

Adams, J.M., Jeppesen, D.G.N., Sanger, F. & Barrell, B.G. (1969). Nature, 223, 1009.

Armstrong, A., Hagopian, H., Ingram, V.M., Sjoquist, I. & Sjoquist, J. (1964). Biochemistry, 3, 1194.

Bawden, F.C. & Pirie, N.W. (1940). Biochem. J., 34, 1278.

Billeter, M.M., Dahlberg, J.E., Goodman, H.M., Hindley, J. & Weissmann, C. (1969). Nature, 224, 1083.

Boedtker, H. & Simmons, N.S. (1957). J.A.C.S., 80, 2250.

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

Brownlee, G.G. & Sanger, F. (1969), European J. Biochem., 11, 395.

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

Caspar, D.L.D. (1963). Advan. Protein Chem., 18, 36.

Cohen, S.S. & Stanley, W.M. (1942). J. Biol. Chem., 144, 589.

Commoner, B., Shearer, G.B., & Yamada, M. (1962). Proc. Natl. Acad. Sci., 48, 1788.

Dube, S.K., Marcher, K.A., Clark, B.F.C., & Cory, S. (1968). Nature, 218, 232.

Egami, F. & Takahashi, E. (1961). J. Biochem. (Tokyo), 49, 1.

Ehresmann, C., Fellner, P., & Ebel, J.P. (1970). Nature, 227, 1321.

Fellner, P., Ehresmann, C., & Ebel, J.P. (1970). Nature, 225, 26.

Fiers, W., Van Montagu, M., De Wachter, R., Haegeman, G., Min Jou, W., Messens, E., Remaut, E., Vandenberghe, A., & Van Styvendaele, B. (1969). Cold Spring Harbor Symp. Quant. Biol., 34, 697.

0 0 0 0 3 8 0 0 3 2 3

Fraenkel-Conrat, H. (1956). J.A.C.A. 78, 882.

Fraenkel-Conrat, H. (1957). Virology 4, 1.

Fraenkel-Conrat, H., Molecular Basis of Virology, N.Y., Reinhold, (1968), 44.

Gierer, A. & Schramm, G. (1956). Z.f. Naturforsch., 11b, 138.

Hall, C.E. (1957). J.A.C.S. 80, 2256

Hart, R. (1958). Biochem. Biophys. Acta, 28, 457.

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

Kado, C.I. & Knight, C.A. (1966). Proc. Nat. Acad. Sci., 55, 1276.

Kado, C.I. & Knight, C.A. (1968). J. Mol. Biol., 36, 15.

Klug, A. & Caspar, D.L.D. (1960). Advan. Virus Res., 7, 225.

Lloyd, D.A. (1969). Ph.D. Thesis, U.C., Berkeley.

Lloyd, D.A. & Mandeles, S. (1970). Biochemistry, 9, 932.

Lowry, O.H., Rosebrough, N.J., Farr, A.L., & Randall, R.J. (1951). J. Biol. Chem., 193, 265.

Mandeles, S. (1967). J. Biol. Chem., 242, 3103.

Mandeles, S. (1968). J. Biol. Chem., 243, 3671.

Mandeles, S. & Bruening, G. (1968). Biochem. Prep., 12, 111.

Mandeles, S. & Tinoco, I., Jr. (1963). Biopolymer, 1, 183.

May, D.S. & Knight, C.A. (1965). Virology, 25, 502.

Mundry, K.W. (1969). Mol. Gen. Genetics, 105, 361.

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

Ohno, T., Nozu, Y., & Okada, Y. (1971). Virology, 44, 510.

Peacock, A.C. & Dingman, C.W. (1967). Biochemistry, 6, 1818.

RajBhandary, U.L., Chang, S.H., Stuart, A., Faulkner, R.D., Hoskinson, R.M., & Khorana, H.G. (1967). Proc. Natl. Acad. Sci., 57, 751.

RajBhandary, U.L. & Stuart, A. (1966). Annu. Rev. Biochem., 35, 759.

Razzell, W.E. & Khorana, H.G. (1961). J. Biol. Chem., 236, 1144.

Rushizky, G.W., Bartos, E.M., & Sober, H.A. (1964). Biochemistry, 3, 626.

Rushizky, G.W. & Sober, H.A. (1962). J. Biol. Chem., 237, 834.

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

Schramm, G., Schumacher, G., & Zillig, W. (1955). Z. Naturforsch., 10b, 481.

Singer, B. & Fraenkel-Conrat, H. (1961). Virology, 14, 59.

Singer, M.F., Hilmoe, R.J., & Heppel, L.A. (1958). Federation Proc., 17, 312.

Stanley, W.M. (1935). Science, 81, 644.

Stanley, W.M. & Lauffer, M.A. (1939). Science, 89, 345.

Steere, R.L. (1963). Science, 140, 1089.

Steinschneider, A. & Fraenkel-Conrat, H. (1966). Biochemistry, 5, 2735.

Steitz, J.A. (1969). Nature, 224, 957.

Stussi, C., Lebeurier, G., & Hirth, L. (1969). Virology, 38, 16.

Sugiyama, T. & Fraenkel-Conrat, H. (1963). Biochemistry, 2, 332.

Symington, J. & Commoner, B. (1967). Virology 32, 1.

Symington, J., Commoner, B., & Yamada, M. (1962). Proc. Natl. Acad. Sci., 48, 1675.

William, R.C. & Steere, R.L. (1951). J.A.C.S., 73, 2057.

Yarwood, C.E. & Fulton, R.W. (1967). Methods in Virology, 1, 237.

LEGAL NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Atomic Energy Commission, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

TECHNICAL INFORMATION DIVISION
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720