Title
PARTIAL NUCLEOTIDE SEQUENCES FOR THREE UNIQUE T-I RNASE FRAGMENTS OF TMV-RNA

Author
Lloyd, David A.

Publication Date
1969

UCRL-18672

ey. 2

PARTIAL NUCLEOTIDE SEQUENCES FOR THREE UNIQUE T-1 RNASE FRAGMENTS OF TMV-RNA

David A. Lloyd

(Ph. D. Thesis)

January 1969

LAWRENCE RADIATION LABORATORY
UNIVERSITY of CALIFORNIA BERKELEY

UCRL-18672

ey. 2

DISCLAIMER

UCRL-18672

UNIVERSITY OF CALIFORNIA

Lawrence Radiation Laboratory
Berkeley, California

AEC Contract No. W-7405-eng-48

PARTIAL NUCLEOTIDE SEQUENCES FOR THREE UNIQUE T-1 RNASE FRAGMENTS OF TMV-RNA

David A. Lloyd
(Ph. D. Thesis)

January 1969

Partial Nucleotide Sequences for Three Unique T-1 RNase Fragments of TMV-RNA

David A. Lloyd

ABSTRACT

Three fragments, having chainlengths of 26, 26, and 70 nucleotides, have been isolated from T-1 ribonuclease digests of TMV-RNA. These fragments appear to be unique in the TMV-RNA molecule. When digested with pancreatic ribonuclease, the three fragments were found to yield the following products:

12 Up, 3 ApUp, 4 $(Ap)_2Up$, 1 Cp, 6 ApCp, 6 $(Ap)_2Cp$, 2 $(Ap)_3$ Cp, 1 Gp;

2 Up, 2 ApUp, 1 $(Ap)_2Up$, 1 $(Ap)_3Up$, 4 Cp, 2 ApCp, 1 $(Ap)_3Cp$, 1 Gp;

5 Up, 2 ApUp, 1 $(Ap)_2Up$, 1 $(Ap)_3Up$, 2 Cp, 2 ApCp, 1 $(Ap)_2Cp$, 1 Gp.

None of the three fragments can be a portion of the gene which codes for the TMV coat protein. Experiments performed by Mandeles[1] indicate that the 70-mer and one of the 26-mers are located extremely close to the 3' end of the viral RNA while the remaining 26-mer is close to the middle of the RNA chain. The similarity of the pancreatic ribonuclease

digestion products of the two 26-mers is discussed with respect to the hypothesis of gene doubling. Attempts to gain further information about the sequences of these fragments are also discussed. In this connection, a possible C-specific nuclease first reported by Anderson and Carter[2] has been investigated.

Two appendices are included. The first of these discusses a theory of ion-exchange chromatography and its utility in predicting the behavior of chromatographic columns. The second appendix describes a method for determining the number of statistically significant component shapes contained in a set of spectra. The method is discussed as a means of obtaining information about the conformation of polynucleotides from their optical properties.

1. Mandeles, S., J. Biol. Chem., 243, 3671 (1968).
2. Anderson, J. H. and Carter, C. E., Biochemistry, 4, 1102 (1965).

ACKNOWLEDGEMENTS

I would like to thank the many members of the Chemistry Department, faculty, fellow graduate students, and non-academic staff, who have given me assistance during my tenure as a graduate student. There are several people to whom I am particularly indebted. The foremost among these is my research advisor, Dr. I. Tinoco, Jr. His guidance and encouragement have been extremely valuable. I am deeply indebted to Dr. Stanley Mandeles who has been the source of a never ending supply of tobacco leaves. The large amount of chromatographic lore which he has imparted to me has been even more valuable than his generous gifts of virus and RNA. Mrs. Barbara Dengler has provided technical assitance of a caliber far above average. Without her help much of this work could not have been completed.

I have benefited from numerous discussions with the faculty members of the biophysical group. Dr. John Hearst, Dr. Robert Harris, and Dr. Jim Wang have all been both friendly and helpful. My association with the other members of Dr. Tinoco's research group has been pleasant and rewarding. The list of people with whom I have worked in close association during the course of my graduate studies includes: Dr. Myron Warshaw, Dr. Charles Cantor, Dr. Daniel Glaubiger, Dr. Richard Jaskunas, Dr. Robert Davis, Dr. Barrett Tomlinson, Dr. David McMullen, Dr. Martin Itzkowitz, Mr. Dana Carroll, Mr. Carl Formoso, Dr. Sunil

Podder, Mr. Mitchell Berman, Mr. Philip Borer, Mr. Mark Levine, Miss Arlene Blum, Dr. Donald Gray, and Dr. Kashy Javaherian.

I would also like to express my appreciation to Mrs. Sharon McConnell and Mrs. Nancy Monroe who helped in the preparation of this dissertation.

TABLE OF CONTENTS

Page

LIST OF ABBREVIATIONS

Chemical

Tris-Cl	(a buffer) 2-amino-2-(hydroxymethyl)-1,3-propanediol adjusted to the given pH with HCl (Sigma)
Na-EDTA	(a buffer and chelating agent) ethylene-diaminetetraacetic acid adjusted to the given pH with NaOH (Eastman)
SDS	(a detergent) sodium dodecyl sulfate (Matheson)
DEAE-Sephadex	(an anion exchange resin) a cross-linked dextran polymer containing diethylaminoethyl groups (Pharmacia)

Optical

ORD	optical rotatory dispersion
CD	circular dichroism
$A_{260}^{cm^2}$	an O.D. unit - the amount of material which, when dissolved in one ml., will produce an absorbance of 1.0 in a one cm. pathlength cell at 260 mμ
$[\Phi]$	specific molar rotation units: (deg.-l./mole-cm.) times 100
λ	wavelength in milli-microns

Nucleic Acids

A	adenosine (a purine nucleoside)
G	guanosine (a purine nucleoside)
C	cytidine (a pyrimidine nucleoside)
U	uridine (a pyrimidine nucleoside)
N	any nucleoside
Np	a nucleoside phosphate (the phosphate is assumed to be in the 3' position unless otherwise stated)

pN	a nucleoside 5' phosphate
N>p	a nucleoside 2',3' cyclic phosphate
poly N	a homopolymer of 3',5' linked nucleoside phosphates
RNA	ribonucleic acid
TMV	tobacco mosaic virus
RNase	ribonuclease

Amino Acids

Ala	alanine
Arg	arginine
Asp	aspartic acid
$AspNH_2$	asparagine
CysSH	cysteine
Glu	glutamic acid
$GluNH_2$	glutamine
Gly	glycine
His	histidine
Ilu	isoleucine
Leu	leucine
Lys	lysine
Met	methionine
Phe	phenylalanine
Pro	proline
Ser	serine
Thr	threonine
Try	tryptophan
Tyr	tyrosine
Val	valine

I. INTRODUCTION

In recent years the sequences of a number of transfer RNA's have been determined.[1,2,3] Sanger and Brownlee[4] have also reported a sequence for the 5S component of ribosomal RNA from E. coli. However, thus far, no sequences have been reported for any RNA which acts as a template for protein synthesis. One of the deterrents to the sequencing of messenger RNA is that, before a sequence determination can be performed, a single messenger RNA must be isolated from a multitude of other messengers, many of which will have nearly identical chainlengths and base compositions. Thus, in the near future it will probably not be feasible to isolate pure messenger RNA's except in unusual cases where their physical and chemical properties are somewhat unique. For example, the extremely short messenger for gramicidin S has been isolated by Hall and Sedat.[5,6] Fortunately, one is not confined to the cellular messenger RNA. One can, instead, work with the viral RNA's. Although the shortest viral RNA's are several times larger than the average messenger RNA, their isolation is straightforward.

The sequencing of even a single gene of a viral RNA is a task of tremendous proportions with the techniques that are now available. There are a number of reasons for undertaking such a task despite the difficulties involved. The comparison of the nucleotide sequence of an RNA message

with the amino acid sequence of the protein produced by this message is the most obvious and most direct way of confirming or correcting the currently accepted form of the genetic code. This form of the genetic code is highly degenerate. Unless the cell has some means of resolving this degeneracy a number of code words could be used interchangeably for a single amino acid. If this were the case one would expect to find mutations of a given species which were phenotypically identical. Thus, a particular strain of a virus might possess individual members whose RNA sequences were quite different. Such a situation would become apparent as one attempted to determine the sequence of the RNA. The sequence of a viral RNA would also increase our understanding of the mechanism which allowed complex organisms to evolve from simple forms of life. For instance, one is interested in looking for homologous sequences in different genes which might arise as a result of gene doubling. Finally, a knowledge of the sequence of an RNA together with information obtained from genetic experiments would allow one to determine the structure of those sites on the RNA responsible for the control of replication and protein synthesis. Such information is necessary for an understanding of genetic control mechanisms at a molecular level.

Attempts to sequence viral RNA have primarily been directed towards either TMV or one of the three closely related Coli viruses MS2, f2, and R17. The five terminal

bases of TMV-RNA have been independently determined by Mandeles[7] and Fraenkel-Conrat[8] using different techniques. A terminal sequence of eleven bases has been determined for MS2.[9] The work to be described here is closely related to the sequencing work on TMV-RNA being performed by Mandeles.

TMV is a rod-shaped virus 3000 Å in length and 180 Å in diameter. The virus has a relatively simple structure which contains only a single strand of RNA and a coat consisting of some 2200 protein molecules. The coat protein molecules appear to all be identical. The RNA is wound in a loose helix containing 45 nucleotides per turn with the molecules of the coat protein packed between the turns. The RNA appears to fit into a groove in the protein. Thus, in the intact virus it is protected from the action of nucleases and other chemical agents.

In the course of his sequence studies on TMV-RNA Mandeles noticed the presence of a long chainlength component in T-1 ribonuclease digests of the RNA.[10] Chromatographic evidence indicated that this component might contain a unqiue fragment of TMV-RNA. The chainlength of this fragment was initially estimated to be about 20 nucleotides on the basis of its elution position in column chromatography. Shortly afterwards, Mundry[11] measured a chainlength of about 40 nucleotides for what appears to be the same fragment. The work described here sets the chainlength of this fragment at 70 nucleotides. In addition two other,

similarly produced, unique fragments of TMV-RNA have been studied.

The study of these fragments was begun with the idea of using them as model compounds for the study of the tertiary structure of nucleic acids. Although some evidence is presented which indicates that the longest of these fragments may be of interest in this connection, the majority of the work is concerned with sequence studies. Interest in this phase of the work was greatly increased by the results of certain experiments performed by Mandeles.[12] The description of these experiments is best postponed until Chapter VI.

II. PREPARATION OF UNIQUE OLIGOMERS FROM TMV-RNA

1. Techniques

Much of the work to be described relies heavily on chromatographic techniques. Therefore, in an attempt to simplify the presentation which is to follow, certain general procedures will be described beforehand. An excellent presentation of the theory of column chromatography, which was useful for understanding column behavior, is given by Vermeulen et al.[13,14] A simple extension of Vermeulen's theory to include gradient elution is given in Appendix I.

The ion-exchange material used was, in all cases, DEAE Sephadex A-25 (Pharmacia). This material requires little washing and gives an eluent free of UV-absorption in the 260 mμ region, which is not true of the styrene resins. More important, it seems to give rise to no adsorptive effects which complicate the prediction of the ion-exchange behavior. These adsorptive effects are present in the sytrene resins and to some extent in DEAE cellulose.[15] For large columns the coarse particle size (50 - 140μ) was employed. However, for small columns (with diameters less than 0.5 cm) the coarse resin was found to give poor resolution and the medium particle size (40 - 120μ was employed. The packing of the coarse resin in columns of small diameter was noticeably

poorer. One could visually detect spaces between the coarse resin particles. This was not true of the medium grade resin. Since the particle sizes listed for the dry forms of the two resins are quite similar, the reason for the difference in packing is not completely clear. It may be that the coarse resin particles swell to a greater extent in water; or the swollen particles of the coarse resin may be more irregular in shape than those of the medium grade resin.

The DEAE Sephadex was prepared for use by packing the desired quantity in a column and passing 10 bed volumes of 2.0 M NaCl, 0.1 M HCl through the column. The resin was then removed from the column and equilibrated with either distilled water or 7 M urea by repeated decantation. This latter process was also used to remove fines. To pack the column, the resin slurry was added in small amounts with continuous stirring. All columns were packed at a salt concentration about 0.1 M greater than the initial gradient concentration, and at a flow rate 5 - 10 times greater than that used for elution. The packed column was then equilibrated with the initial salt concentration of the elution gradient before loading the sample. This procedure resulted in a tightly but evenly packed column, and reduced the bed shrinkage caused by increasing salt concentration.

All columns were run at room temperature as a matter

of convenience. There was no indication of degradation of RNA during the chromatography. Precise flow rates and elution conditions will be given later; but in general, flow rates were in the range of 12 - 25 ml/hr cm^2. Salt concentration gradients were 0.008 - 0.010 Moles/unit eluent volume, where the unit of eluent volume is equal to the total column volume.

A comment is in order concerning the treatment of glassware used in this work. Contamination by the ribonucleases can occur from a number of sources[16] and is difficult to remove. It has been observed[17] that nucleases remain bound to glass surfaces and retain their biological activity even after treatment with detergent solutions. Thus, all glassware used in this work was washed in hot detergent and/or chromic acid, thoroughly rinsed, dryed, and, after being covered with aluminium foil, baked for 12 hours at 180 - 200°C.

Due to the danger of nuclease contamination of buffers by even small amounts of bacterial growth, all buffers capable of supporting the growth of bacteria were stored at 4°C and prepared fresh weekly from concentrated stock solutions of the acidic or basic salt. Although anionic clays such as bentonite and Macaloid have been shown to be extremely effective at inhibiting nuclease activity[18], they were not employed except as specifically mentioned in the early preparative stage of the work. It was found that these materials contributed spurious UV absorption

to the samples, which complicated the identification and quantitative assay of oligonucleotides by UV absorption spectroscopy. Unless otherwise stated, all chemicals were reagent grade from either Baker and Adamson or Mallinckrodt.

2. Preparation and Digestion of TMV-RNA

Tobacco Masaic Virus (TMV) was isolated from infected tobacco leaves by homogenization of the leaves followed by differential centrifugation of the leaf extract. The viral ribonucleic acid (RNA) was separated from the coat protein by the phenol method in the presence of Macaloid (an anionic clay) to prevent degradation by extraneous nucleases. The infected leaves were supplied by Dr. Stanley Mandeles of the Space Sciences Laboratory, University of California, Berkeley. Both the procedure and the materials used therein are described in detail by Mandeles and Bruening.[19] The yields obtained from the above procedure as well as from the steps yet to be described are summarized in Table I of Chapter II.3.

TMV-RNA, in 200 mg batches, prepared as described above, was hydrolyzed with ribonuclease T-1 (Cal Biochem) in a Radiometer TTT-1 Titrimeter at 40°C. The RNA, as the ethanol precipitate, was dissolved in distilled water to give a solution containing 3 - 4 mg/ml RNA. This solution was centrifuged at 17,000 x g, 4°C, for 30 minutes to remove the Macaloid introduced in the RNA preparation.

The supernatant was collected and recentrifuged as above until the resulting supernatant was free of turbidity. Due to the high viscosity of the RNA solution, as many as three centrifugations were sometimes necessary to remove the last traces of Macaloid. Incomplete removal of the Macaloid leads to inactivation of ribonuclease and incomplete hydrolysis of the RNA.

The RNA solution was brought to pH 7.8 by the addition of 0.5 M KOH and allowed to stand in the titrimeter at 40°C for 30 minutes, before the addition of enzyme, to insure that no spontaneous hydrolysis was occurring. A total of 15,000 units* of ribonuclease dissolved in 0.8 ml distilled water was used. Two-thirds of the enzyme was added initially and the remainder was added two hours later. The pH was maintained at 7.8 by the addition of 0.5 M KOH until the reaction reached completion after 6 - 8 hours. The consumption of KOH was routinely found to be 92 - 97% of the value predicted from the UV absorption of the RNA. The hydrolysate was brought to 7 M in urea by the addition of solid urea and then was applied to the top of a chromatographic column packed with DEAE Sephadex preparatory to

---

*The unit of T-1 ribonuclease activity has been defined by Takahashi.[20] A 1.0 ml reaction mixture containing enzyme, 0.75 mg yeast RNA, 50 μM pH 7.5 Tris-Cl buffer, and 2 μM EDTA is incubated for 15 min. at 37°C. The reaction is terminated by the addition of 0.25 ml of 0.75% uranylacetate, 25% $HClO_4$. After centrifugation 0.2 ml of the supernatant is diluted to 5.0 ml. One unit of T-1 ribonuclease will increase the 260 mμ absorption of the diluted supernatant by one absorbance unit.

performing a chainlength fractionation of the fragments resulting from RNA digestion.

3. Chromatographic Isolation of Unique Oligomers

The DEAE-Sephadex column used for the chainlength separation was packed in a 100 ml burette, allowing 1 ml of bed volume per 2 mg of digested RNA, and washed with 300 ml of 7 M urea. Better resolution can be obtained at loading ratios of 0.5 - 1.0 mg RNA per ml bed volume; but the higher load allowed greater ease in handling the amount of RNA involved.

The elution of the T-1 fragments was accomplished with a linear concentration gradient from 0.1 M to 0.6 M sodium chloride in 5 liters of 7 M urea. Because of the low flow rate used, 40 ml per hour, a period of 4 - 5 days was required for complete elution. However, a greater flow rate, i.e., 100 ml/hr, gave a 50% decrease in the yield of the longest T-1 fragment. The decrease in yield can be attributed to the fact that at large flow rates the eluent does not have time to equilibrate with the adjacent resin particles. This departure from equilibrium causes the sample to be eluted in a broad band with long tails. Thus a smaller percent of the sample is collected in the peak fractions. This effect is most pronounced with long fragments which, due to their low diffusion coefficients, require a longer period of time to reach equilibrium.[14]

The eluent fractions labeled psi and omega in the elution profile shown in Fig. 1 contained approximately 1.5 mg and 2.0 mg of oligonucleotide, respectively. They were collected separately for concentration and desalination. The collected fractions were first diluted five-fold with distilled water and then passed through a DEAE-Sephadex column, having a bed volume of 1.0 ml (0.35 × 10 cm), at a flow rate of 50 ml/hr. The long T-1 oligomers remained bound on the column, which was then washed with 5 ml 7 M urea. The long oligomers were eluted in a volume of 1 - 2 ml by passing 1.0 M NaCl, 7 M urea through the column. Upon the addition of four volumes of ethanol to the column eluent, at 0°C, the oligonucleotides precipitated while the NaCl and urea remained in solution. (KCl will precipitate under these conditions.)

After standing for at least 4 hours at 0°C, the precipitates were collected by centrifugation in a clinical centrifuge, washed with 95% ethanol, dried for 30 minutes under vacuum, and stored at -20°C.

The above procedure for concentration and desalination was found to be far superior to successive dialyses and lyophilizations, in that it gave a high yield of material (95 - 97%) with no evidence of degradation and could be performed in a shorter period of time. In addition the oligomers were obtained in a more convenient form and the shapes of their UV absorption spectra indicated that they

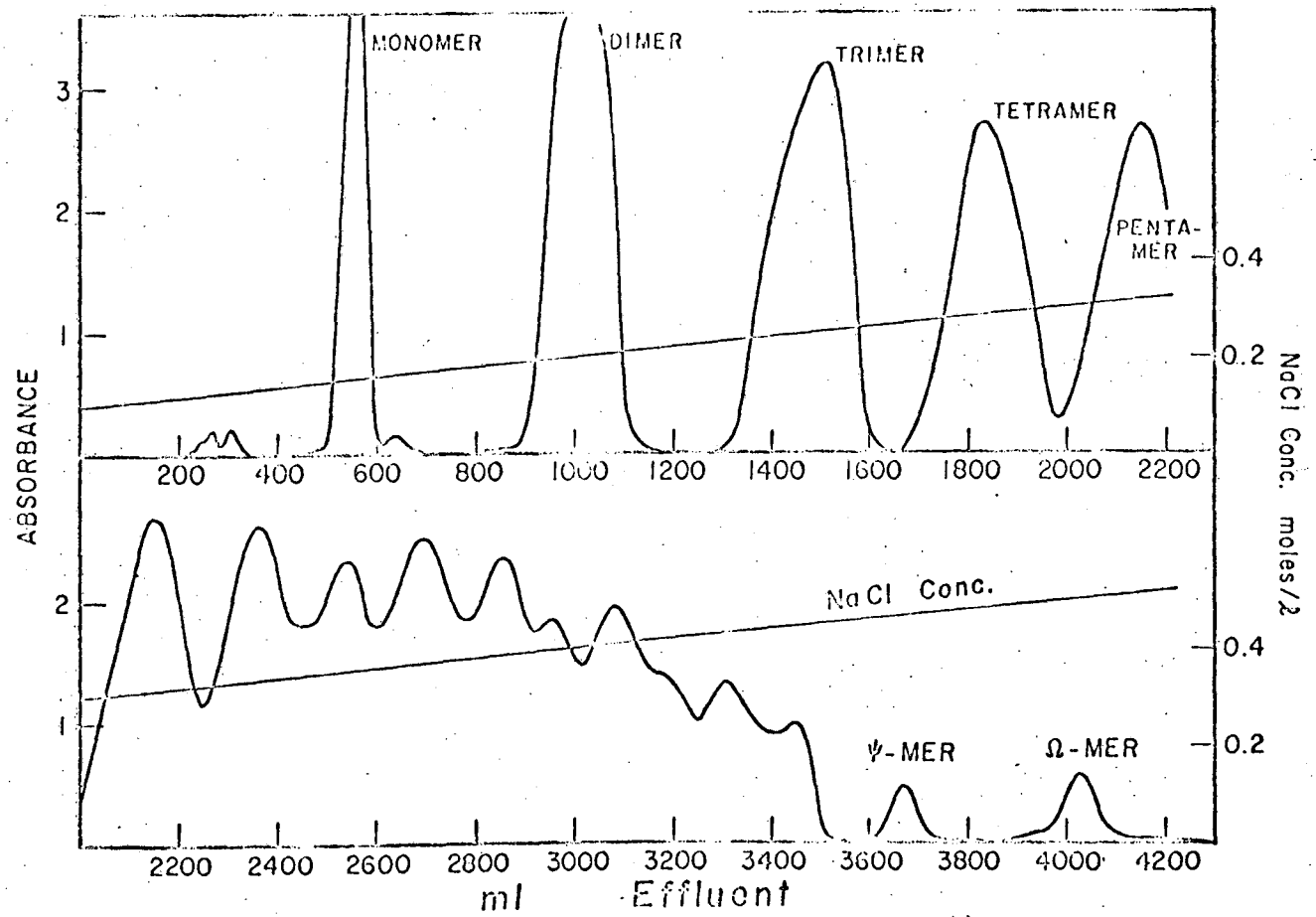


Figure 1. Chainlength separation of T-1 oligomers of TMV-RNA.

Resin: DEAE-Sephadex, A-25
Solvent: 7M Urea, pH 7.5
Bed Volume: 100 ml
Sample Load: 200 mg
Flowrate: 40 ml/hr

contained no non-nucleotide UV absorbing material which could interfere with later chromatographic results.

Immediately before use, the psi and omega fractions were further purified by rechromatography. A column 0.4 cm × 21 cm was packed with DEAE Sephadex which had been washed as previously described. However, in this case the medium particle size was employed, as it was observed that the coarse resin gave poor resolution when used in columns of small diameter. A sample, containing 1.5 - 2.0 mg of the psi-mer or omega-mer, was dissolved in 2 ml 7 M urea and applied to the top of the column. The elution was done at a flow rate of 2 ml/hr with a linear concentration gradient of 0.25 M NaCl to 1.0 M NaCl in 250 ml of 7 M urea. The results of these chromatograms, which are shown in Figs. 2 and 3, demonstrate that each of the preparations involved contains either a single oligomer or a number of oligomers which are nearly identical in chainlength. It was possible to resolve the rechromatographed psi-mer fraction into the two component labeled psi-1 and psi-2 in Fig. 4. This was accomplished by chromatography on DEAE Sephadex at pH 2.8 in 7 M urea. Under these conditions the separation obtained is due to the difference in the uridine content of the two oligomers.

In contrast to the psi-fraction, the omega-fraction shows a single component upon chromatography at pH 2.8 in 7 M urea. The low pH chromatography also showed the

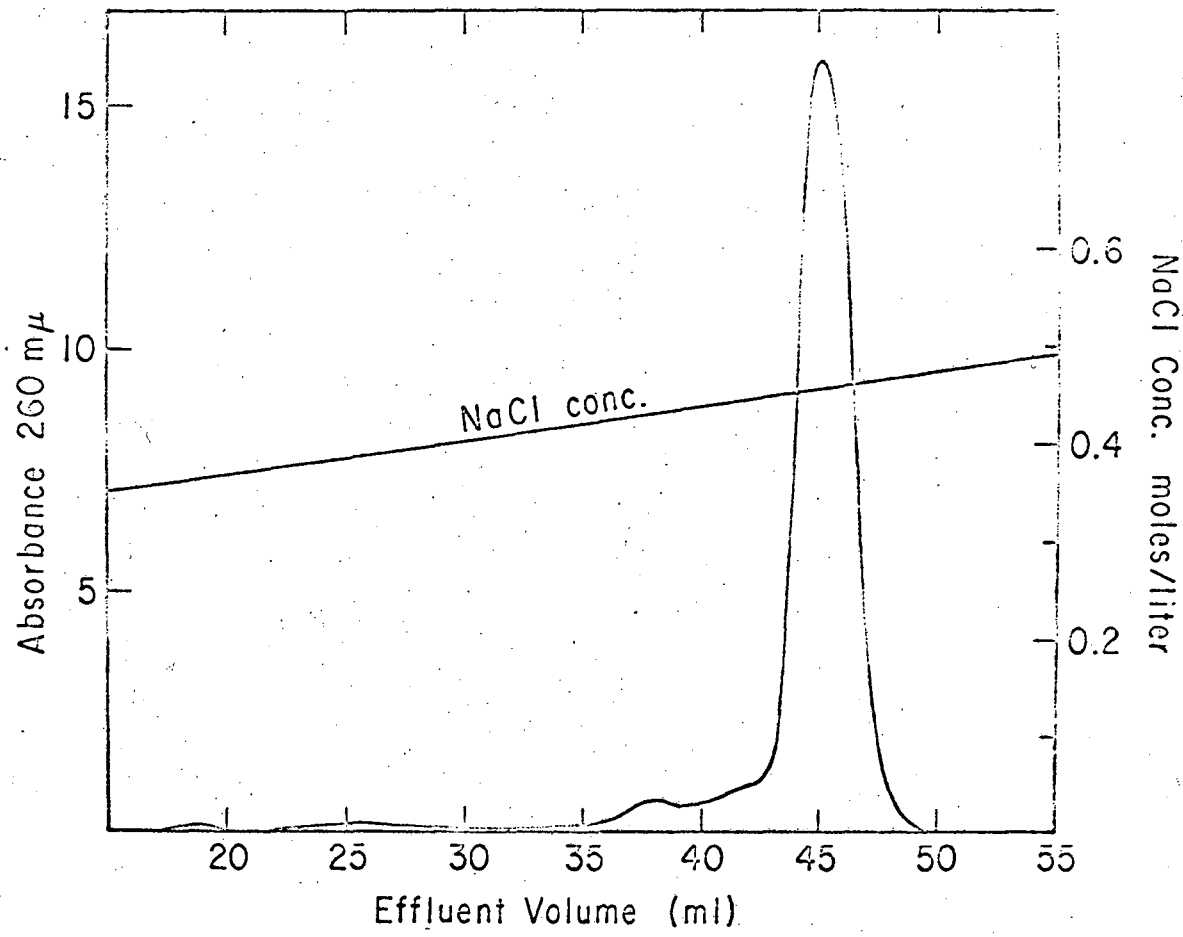


Figure 2. Rechromatography of the omega-mer.

Resin: DEAE-Sephadex, A-25

Solvent: 7 M Urea, pH 7.5

Bed Volume: 3 ml

Sample Load: 2.0 mg

Flowrate: 2 ml/hr

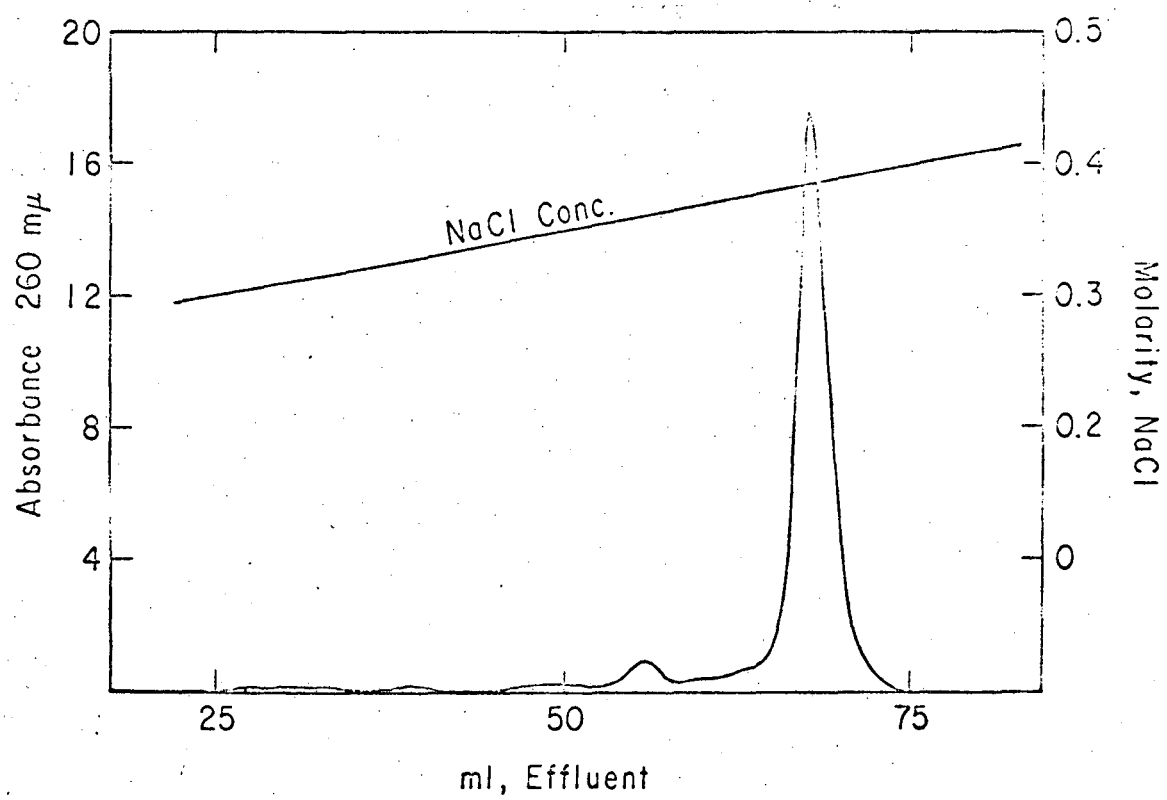


Figure 3. Rechromatography of the psi-mers.

Resin: DEAE-Sephadex, A-25

Solvent: 7 M Urea, pH 7.5

Bed Volume: 3 ml

Sample Load: 1.5 mg

Flowrate: 2 ml/hr.

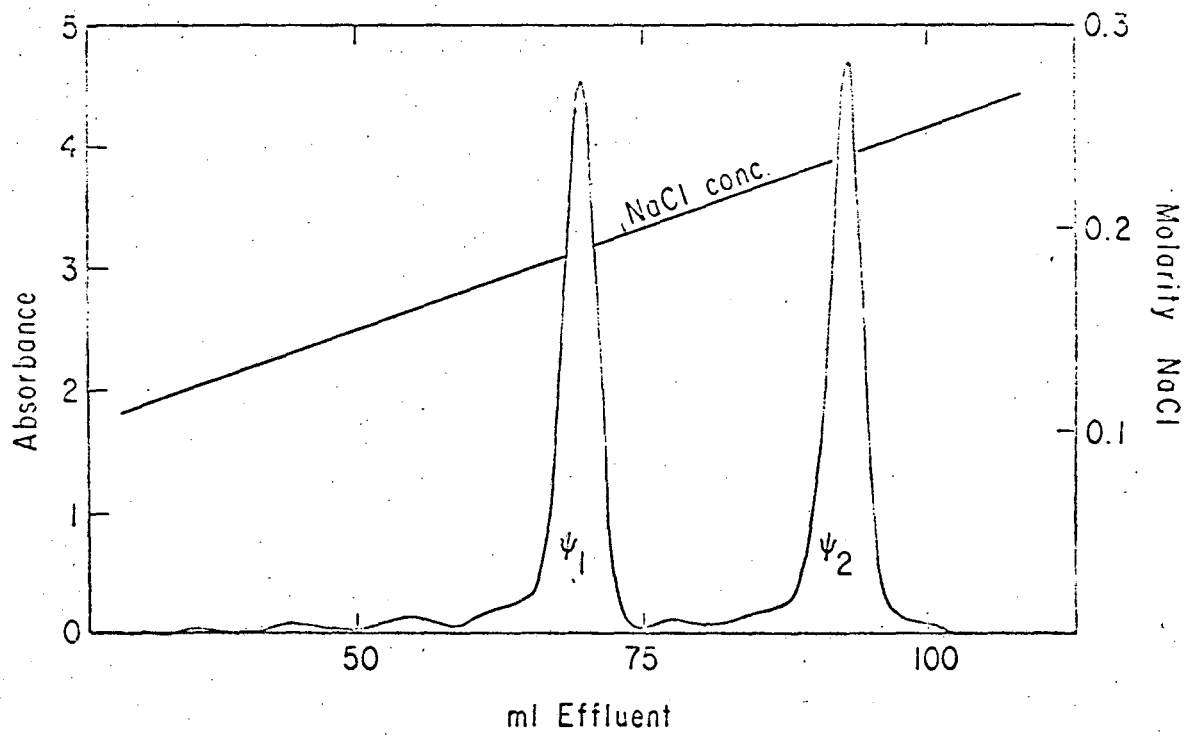


Figure 4. Separation of the psi-mers.

Resin: DEAE-Sephadex, A-25

Solvent: 7 M Urea, pH 2.8 (HCl)

Bed Volume: 4 ml

Sample Load: 1.0 mg

Flowrate: 1.8 ml/hr

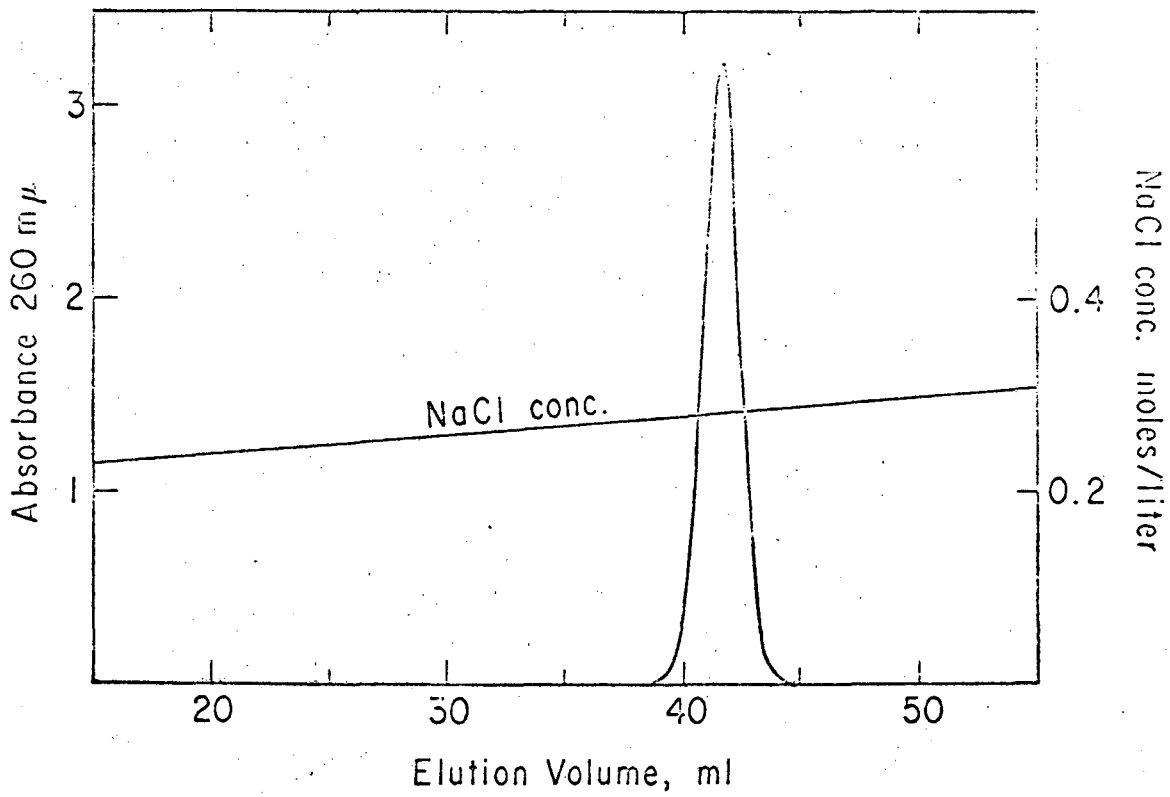


Figure 5. Rechromatography of the omega-mer at low pH.

Resin: DEAE-Sephadex, A-25

Solvent: 7 M Urea, pH 2.8 (HCl)

Bed Volume: 4 ml

Sample Load: 0.2 mg

Flowrate: 1.8 ml/hr

omega-mer preparation to be free of minor contaminants. The elution profile in Fig. 5 shows that outside of the region in which the omega-mer was eluted the absorption of the eluent at 260 mμ was always less than 0.02 absorbance units. Since the low pH chromatography resulted in no further purification of the omega-mer, it was not routinely performed. The yields obtained at various stages in the preparation of the psi-mers and omega-mer are given in Table I. The time required to prepare the amounts of the three oligomers shown in this table, starting from the infected leaves, was about one month.

For the purpose of clarity, it is desirable to attempt to summarize and to draw some conclusions from the information gained in the course of the preparation described above. Three oligomer fractions (omega, psi-1, psi-2) have been isolated from a T-1 RNase hydrolysate of TMV-RNA. One is interested in finding out whether it is possible to estimate the chainlengths of these oligomer fractions from the positions at which they are eluted in the chainlength chromatography. Fig. 6 illustrates an attempt to determine the chainlength of the omega-mer and psi-mers from chromatographic data. The chainlengths associated with the first three peaks of the elution profile in Fig. 1 have been determined to be one, two, and three by comparison of the UV absorption and ORD spectra of the oligomers present in those peaks with the spectra obtained by Warshaw[21] and Cantor.[22] Beyond chainlength three, one can, with confidence,

TABLE I

YIELDS OBTAINED FOR VARIOUS PREPARATIVE PROCEDURES

Amount of Starting Material	Yield of Product
2.4 Kg infected tobacco leaves- - -	2-4 gm TMV
5 gm TMV- - - - - - - - - - - - - - - -	200 mg TMV-RNA
200 mg TMV-RNA- - - - - - - - - - - -	*55 $A_{260}^{cm^2}$ omega-mer
- - - - - - - - - - - -	45 $A_{260}^{cm^2}$ psi-mers
	(yield after rechromatography)
- - - - - - - - - - - -	50 $A_{260}^{cm^2}$ omega-mer
- - - - - - - - - - - -	15 $A_{260}^{cm^2}$ psi-1-mer
- - - - - - - - - - - -	15 $A_{260_1}^{cm^2}$ psi-2-mer

*1.0 mg is approximately equivalent to 30 $A_{260}^{cm^2}$.

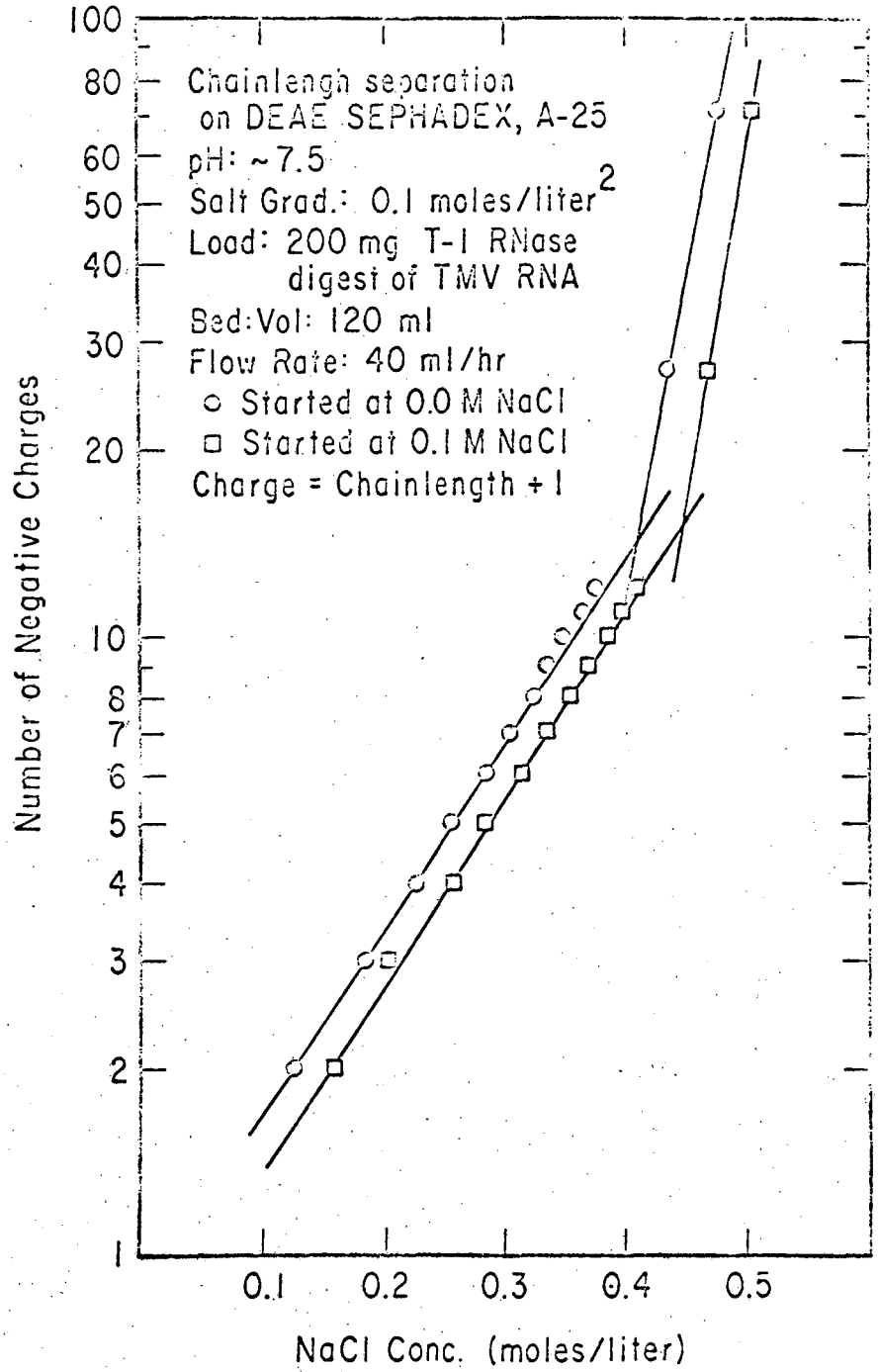


Figure 6. An empirical relationship between the chainlengths of oligonucleotides and their elution volumes in ion-exchange chromatography.

assign consecutive chainlengths to the succeeding peaks up to chainlength seven or eight.[23] At this point the loss of resolution prevents one from continuing this procedure. If one plots the logarithm of the charge on these short oligomers versus the salt concentration at which each oligomer is eluted, one obtains a straight line. This line can be extrapolated to the salt concentrations at which the psi and omega fractions are eluted to obtain respective chainlength estimates of 17 and 23.

The determination of the true chainlengths for the psi-mers and omega-mer will be discussed later. However, if one places the true values (a chainlength of 26 for both psi-mers and a chainlength of 70 for the omega-mer) on the graph in Fig. 6, one sees that a drastic change in the slope of the plot must occur at about chainlength 13 or 14. These chainlengths represent a molecular weight of about 5,000. It is suggestive that Sephadex G-25, which presumedly contains the same amount of crosslinking, has been found to exclude polypeptides and polysaccharides with molecular weights greater than 5,000. It may be that oligomers longer than fourteen nucleotides are incapable of reaching large numbers of ion-exchange sites which are inside the pores of the resin particles. Thus, the effective column capacity would be greatly reduced for these longer oligomers. If this explanation is correct, one would expect chainlength chromatography on Sephadex-A50 to show a linear relationship between the logarithm of the charge

and the elution volume up to a chainlength of approximately 30 nucleotides. An alternative explanation is that the nature of the binding of oligonucleotides to the resin changes as the chainlength increases due to the increased flexibility of the longer molecules. It is difficult to imagine how this explanation could account for such a sharp break in the curve.

The psi-mers and omega-mer all appear homogeneous in the two chromatographic systems utilized in their isolation. At the time these fractions were first isolated, the significance of their chromatographic homogeneity could not be assessed. However, the results of the experiments to be presented in Chapter IV.3 indicate that each of the three oligomer preparations is homogeneous and allow one to estimate the resolution of the chromatographic techniques which have been described in this chapter. One can detect the presence of two oligomers in an elution profile when the two oligomers are separated by more than one bandwidth, where the bandwidth is defined to be the width of a peak at half its maximum value. The bandwidths for the psi and omega peaks in Fig. 1 are 60 ml to 80 ml. One can use the chainlength plot shown in Fig. 6 to relate this increment of eluent volume to a range of chainlengths. For chainlengths on the order of 70 nucleotides one can detect the presence of two oligomers which differ in chainlength by 10 nucleotides. While, for chainlengths on the order of

25 nucleotides, one can resolve chainlength differences as small as ± nucleotides. These results apply to the initial chainlength chromatography. One seems to achieve slightly better resolution under the conditions used for the rechromatography of the psi-mers and omega-mer.

The resolution of the low pH chromatography appears to be roughly similar to that obtained a neutral pH provided that one equates the charge on an oligomer with the uridine content rather than the chainlength. However, the chromatographic separations performed at pH 2.8 resolve oligomers almost entirely on the basis of uridine content. Thus, long oligomers differing only in their adenosine and cytidine contents, would probably not be resolved at this pH. This difficulty could be remedied by performing a third chromatography at pH 3.5. This was not found to be necessary for the isolation of the psi-mers and omega-mer.

III. PHYSICAL STUDIES ON THE OMEGA-MER

1. Chainlength Estimation from Sedimentation Equilibrium

The preceding chapter ended with an attempt at estimation of the chainlengths of the psi-mers and omega-mer from chromatographic elution patterns. It was desirable to have at least a rough estimate of the chainlength obtained by a more reliable technique, especially for the omega-mer, which is the longest of the three fragments. The standard method of chainlength determination for small oligonucleotides is end-group analysis. This technique can be applied in three ways. (1) The amount of phosphate ion liberated by alkaline phosphatase, which removes only the terminal phosphate, can be compared to the total phosphate content of the oligomer. (2) The terminal phosphate can be removed; and, after complete digestion and chromatographic separation, the ratio of nucleotides to nucleoside can be determined. (3) When the oligomer contains a unique nucleotide (in this case the terminal guanidylic acid left after T-1 ribonuclease hydrolysis) the amount of this nucleotide can be compared to the total amount of nucleotide contained in the oligomer. All three of the above variations require the comparison of two experimental results, each of which will contain an error of approximately $\pm$ 2%. Thus, the chainlength obtained will be good to only $\pm$ 4%. It seemed that a chainlength determination having about the same degree of

accuracy could be obtained from a sedimentation equilibrium experiment using UV absorption optics. This method has the advantage of requiring only 0.004 mg of sample as compared with the minimal value of 0.5 mg required by the former methods.

A sedimentation equilibrium run was carried out on the omega-mer at 0°C in 0.1 M NaCl, 0.05 M PO_4, 0.001 M EDTA, pH 7.5 buffer for 76 hours. The rotor speed and column height were of the order of 10,000 rpm and 5 mm respectively. The mean radius was 6.9 cm. Plots of the logarithm of the concentration versus the square of the radius constructed for times between 22 hours and 76 hours gave straight lines, see Figure 7, whose slopes were used to determine the apparent molecular weight as a function of time.

TABLE II

Time (in hours)	Apparent Molecular Weight
22	16,700
31	18,900
46	21,100
53	21,700
76	23,000

The values in Table II were calculated on the basis of a

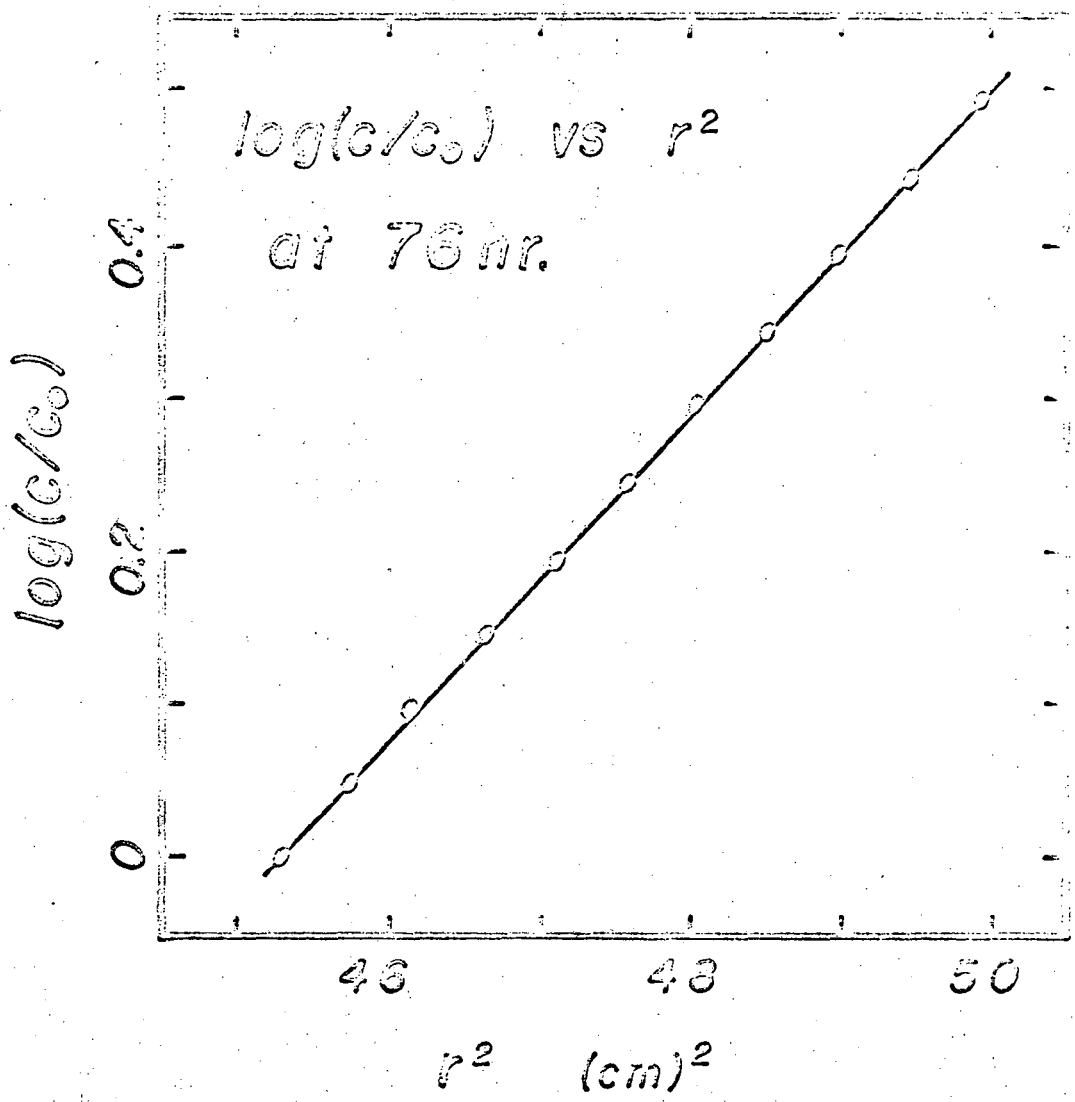


Figure 7. Sedimentation equilibrium distribution for the omega-mer.

Apparent molecular weight at 76 hr is 23,000.

partial specific volume of 0.54 for RNA as measured by Tennent.[24] The solvent density was 1.010 gm/ml.

In order to calculate a chainlength from the molecular weight it is necessary to assume an average molecular weight per nucleotide. At the time molecular weight measurement was made, the base composition of the omega-mer was not known. Thus, the molecular weights of the three nucleotides Ap, Up, and Cp and their sodium counter ions were averaged with equal weight. The nucleotide Gp was neglected since an oligomer obtained from a T-1 ribonuclease digest should contain a single terminal guanosine residue. A value of 354 was obtained for the per residue molecular weight. This value could be in error by as much as $\pm$ 3%. Taking the molecular weight of the omega-mer and its sodium counter ions as 23,000, one obtains a chainlength of 65 nucleotides. In view of the uncertainty in the base composition and the fact that no correction was made for the charge on the oligomer, this value was assumed to be accurate to only $\pm$ 10 nucleotides.

Sedimentation equilibrium has been shown to give molecular weights to better than 1% accuracy for sucrose and globular proteins such as ribonuclease.[25] However, this author can find no report of its use in determining chainlengths for oligonucleotides. Thus, when the base composition and chainlength (70 nucleotides) were determined by the work described in Chapter IV, there was an opportunity

to check the accuracy of this method of chainlength determination. For this purpose two corrections were applied to the data.

According to Van Holde and Baldwin[25] a plot of the logarithm of the difference between the apparent molecular weight at equilibrium and the apparent molecular weight at some earlier time should be linear with respect to time. Figure 8 shows such a plot for three trial values of the apparent molecular weight at equilibrium. The value of 23,700 was chosen as the best value of the apparent molecular weight at equilibrium.

Because of the high charge possessed by oligonucleotides, the charge correction is significant even at the large salt concentrations used in this experiment. The charge correction was calculated from the following expression:[26]

$$n\,\bar{M} = M_2 = M_{app} + (n+1)(z/2)L/(1-\bar{v}_2\rho)$$

where n is the chainlength, ρ is the solvent density, $\bar{v}_2$ is the partial specific volume of the sodium salt of RNA, and z is the fractional effective charge. $\bar{M}$, M_2, and M_{app} refer to the molecular weights of the monomer, the omega-mer and to the measured apparent molecular weight. L is given by,

$$L = \sum_i x_i M_i (1-\bar{v}_i \rho) \quad ,$$

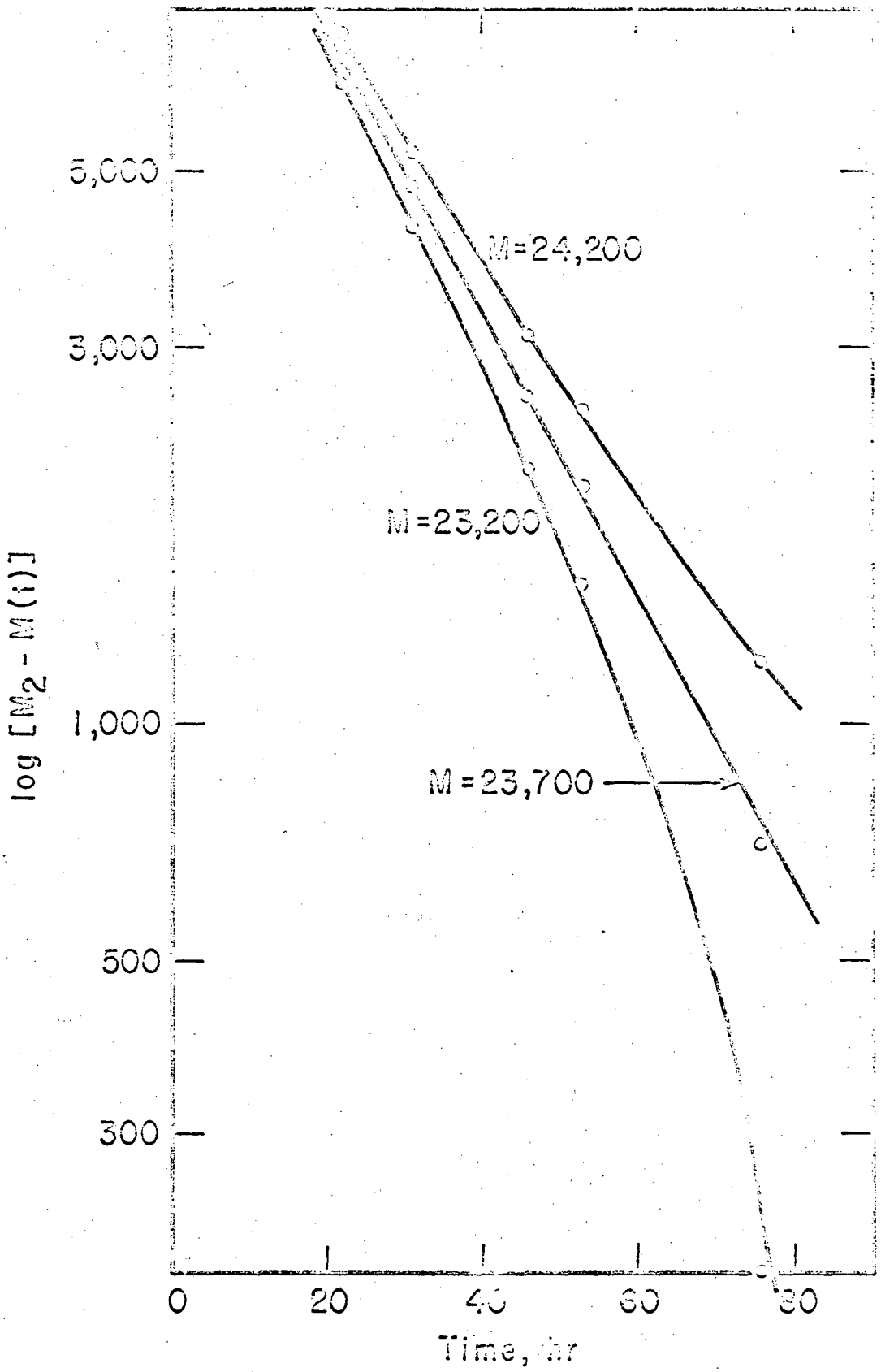


Figure 8. Determination of the apparent molecular weight at infinite time.

where x_i is a mole fraction, M_i is a molecular weight, $\bar{v}_i$ is a partial specific volume, and the subscripts refer to the salt species contained in the supporting electrolyte. Taking L=60, z=0.4,[27] and the average molecular weight of the monomer with its sodium counter ion as 359, the chainlength of the omega-mer was estimated to be 71 $\pm$ 4 nucleotides. Although the charge correction term is only roughly estimated, it is a small correction and probably contributes an uncertainty of only $\pm$ 2% to the final result.

2. Temperature Dependence of the ORD of the Omega-mer

The temperature dependencies of various optical properties have been previously used for studying conformation changes of nucleic acids in solution.[28,29] Theoretical studies of the ORD of RNA[30,31] indicate that this property is quite sensitive to changes in the geometry of polymers of this type. McMullen, Jaskunas, and Tinoco[32] have used the temperature dependence of the ORD to study changes in the conformation of TMV-RNA. They found that they were able to express their data as linear combinations of two basic curve shapes. By comparison of these two shapes with the ORD of model compounds they were able to identify the two component shapes with the single and double strand forms of TMV-RNA.

An investigation of the temperature dependence of the ORD of the omega-mer was initiated to determine whether this oligonucleotide contains intramolecular double stranded regions. If such regions were found to be present, this molecule could be used as a model compound to study hydrogen bonding between two regions of a single strand of RNA. Although the preliminary results which were obtained were very encouraging, the attention given to the sequencing studies described in the succeeding chapters prevented further investigation of the optical properties of the omega-mer. The meager results presented here do not allow one to reach any definite conclusions about the conformation of the omega-mer. However, they indicate that further investigation would be profitable; and they serve as an illustration of how one might apply the method of data analysis presented in Appendix II to the problem of polymer conformation.

Approximately 0.5 mg of omega-mer was dissolved in 1-2 ml of distilled water. This solution was dialyzed against 4 ℓ of 10^{-3} M Na-EDTA, pH 7.5, for three 4-hour periods. The dialysate was diluted with 10^{-3} M Na-EDTA, pH 7.5, to give a 260 mμ absorbance of approximately 0.5. This stock solution was used for all of the ORD measurements. The ORD was measured with a Cary 60 spectropolarimeter equipped with a Datex attachment for digital output on paper tape. The ORD was read at 0.5 mμ intervals between

225 and 350 mμ. The data thus obtained were smoothed with a 10 point quadratic smoothing program[33] which printed out ORD values at 1 mμ intervals. The concentration of the omega-mer solution used for this work was determined from the absorbance at 260 mμ on the basis of an assumed extinction coefficient of 10,400 per mole of nucleoside at 25°C. The concentration was used to obtain the molar rotation per nucleoside. A fresh aliquot of the stock solution was used for each temperature. The sample, in a stoppered 2 cm cuvette, was held within ± 0.2°C of the desired temperature with a Haake circulating bath, type F. Since the density of the solvent varied by only 1% over the experimental temperature range, no correction for solvent expansion was made.

The reader should acquaint himself with the method and notation introduced in Appendix II before reading the discussion which follows. ORD spectra were measured at six temperatures. Four of these spectra are shown in Figure 9. The set of spectra were analyzed by the method described in Appendix II. Since the portions of the spectra at long wavelengths did not show any change in shape with temperature, only the values of the ORD at 1 mμ intervals between 230 and 306 mμ were employed in the analysis. The results which were obtained are shown in Table III. These results indicate the presence

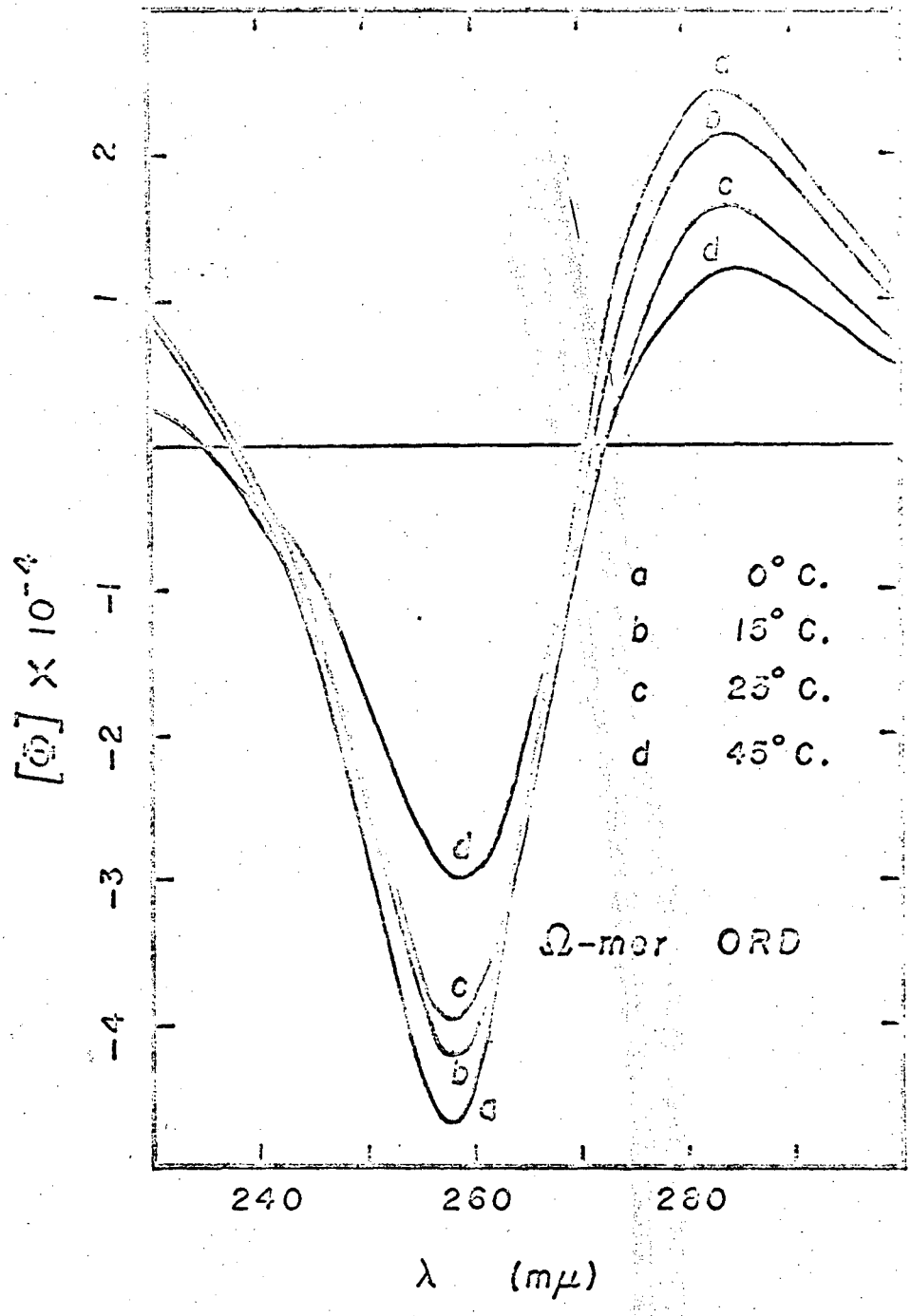


Figure 9. Temperature dependence of the omega-mer ORD in 10^{-3} M Na-EDTA, pH 7.5.

TABLE III

Determination of the Number of Significant Basis Spectra for the Omega-mer ORD Spectra Taken at 0°C., 7°C., 15°C., 25°C., 35°C., and 45°C. in 10^{-3} M Na-EDTA, pH 7.5; m = 77, n = 6

μ	e_μ	degrees of freedom	σ_μ^2	$\sigma_\mu^2/\sigma_{\mu+1}^2$	F(5%)*	F(1%)**
0	1660	462	3.62	110	1.17	1.25
1	11.6	385	0.0329	9.17	1.20	1.29
2	0.757	308	0.00359	2.38	1.25	1.37
3	0.214	231	0.00151	1.72	1.29	1.43
4	0.0907	154	0.000879	1.52	1.40	1.61
5	0.0446	77	0.000579			

*The probability that the ratio of variances will exceed this value is 5%.

**The probability that the ratio of variances will exceed this value is 1%.

of at least two significant components, and possibly a third. The F-test indicates that, if the third component is significant, it is just barely above the level of the experimental error. Thus, this component will be ignored; and the set of data will be treated as a two component system. The possible significance of such a third component, as well as a possible means of treating a three component system, will be discussed later.

The temperature dependence of the optical properties of RNA was originally attributed entirely to the presence of hydrogen bonded regions[34,35] which were thought to possess a conformation similar to the double strand DNA helix. More recently temperature studies on both homopolymers[36,37,38] and dinucleotide phosphates[39,40] have shown that single strand RNA possesses a definite secondary structure. This structure apparently owes its stability to interactions between neighboring bases in the RNA chain. This body of experimental data leads one to believe that the ORD of an RNA can be explained in terms of three classes of conformations. The first type of conformation is the random coil in which no interactions exist between the bases of the RNA. The ORD spectrum of the random coil should be very similar to the sum of the ORD spectra of the constituent monomers. The next structure which must be considered is that of the single strand helix, or stacked form of the polymer. Finally, one must consider the ORD

signal caused by hydrogen bonded, or double stranded regions. Possible triple stranded regions will also be considered in this category. The basic assumption underlying such a model is that the contributions, of these three classes of conformations, to the ORD of an RNA are additive. Thus, the ORD spectrum of an RNA can be treated as a superposition of the ORD spectra of three basic structures. McMullen, Jaskunas, and Tinoco[32] were able to interpret their ORD data from TMV-RNA in terms of such a model. The fact that the above analysis of the ORD spectra of the omega-mer indicates the presence of only two major components suggests that such a model can also be applied to the omega-mer ORD data.

The most obvious indication that the ORD spectrum of the omega-mer changes its shape as a function of temperature is the fact that the long wavelength crossover point shifts to the red as the temperature is increased. The same qualitative behavior was observed for TMV-RNA by McMullen, Jaskunas, and Tinoco. These authors were able to identify the limiting shape obtained at high temperature with the spectrum of the single strand helical conformation. This was done by comparing the shapes of the experimental spectra at high temperatures with the spectrum which Cantor[41] has calculated for the single strand form of TMV-RNA. One would like to make such an identification for the omega-mer ORD spectrum taken at

45°C. At present it is not possible to make this identification in a convincing manner. Although the largest change in shape occurs in the vicinity of 20°C, as will be shown below, one cannot be sure that further changes in shape do not occur above 45°C without examining spectra taken at higher temperatures. Secondly, while Figure 10 shows that the 45°C omega-mer spectrum possesses a qualitative similarity to the high temperature TMV-RNA spectra, there are fairly large quantitative discrepancies between their shapes. Actually, one would not expect the shapes to be identical since the base composition of the omega-mer, as determined in Chapter IV, is quite different than that of TMV-RNA. One would actually like to compare the 45°C omega-mer spectrum against a weighted sum of the dinucleoside phosphate spectra, as was done by McMullen, Jaskunas, and Tinoco. Unfortunately, the ORD of a single strand structure with the base composition of the omega-mer has not been calculated. Since the presence of the bases G and C tends to increase the cross-over wavelength, one would expect the crossover wavelength of such a calculated ORD spectrum to be shifted in the same direction, relative to the TMV-RNA spectrum, as is the crossover wavelength of the 45°C spectrum.[41]

The analysis of the change in shape of the omega-mer ORD will be continued on the assumption that the 45°C spectrum

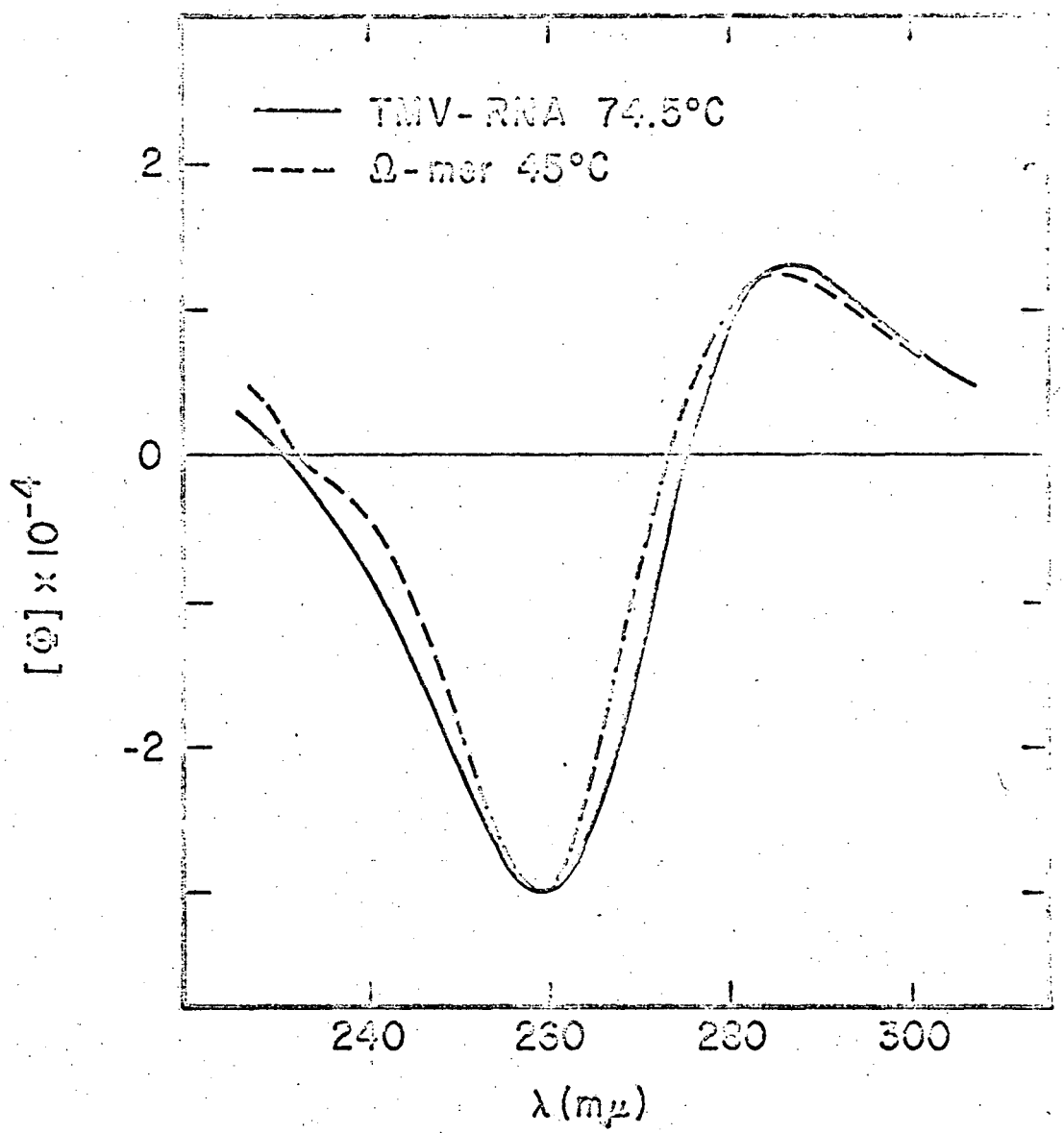


Figure 10. A comparison of the ORD of the omega-mer at 45°C with the ORD of TMV-RNA at 74.5°C.[32]

The curve for TMV-RNA was normalized to the omega-mer curve at 260 millimicrons.

can be associated with the ORD of the single strand conformation. For the present this assumption must remain somewhat speculative. The ORD which arises from the random coil conformation (the ORD of the monomers) has been found to be negligible compared to the spectra of the single and double strand conformations.[32] If the 45°C spectrum can be identified with the ORD of the single strand conformation, then the change in the ORD of the omega-mer at the crossover wavelength (272.6 mμ) of the 45°C spectrum will be the melting curve for the double strand conformation. This melting curve is shown in Figure 11. A melting curve obtained in this manner may be considerably in error since the points in curve depend upon the value of the ORD measured at a single wavelength. It is possible to use the eigenvectors, $\underset{\sim}{U}_\alpha$, and the eigenvalues, e_α, defined in Appendix II, to calculate the same melting curve in a manner such that each point in the melting curve is an average, over all wavelengths contained in the spectrum, corresponding to that point.

When the set of spectra were analyzed and found to contain two significant components, a pair of orthonormal basis vectors, $\underset{\sim}{V}_1$ and $\underset{\sim}{V}_2$ were also determined. As was shown in Appendix II these basis vectors can be written as

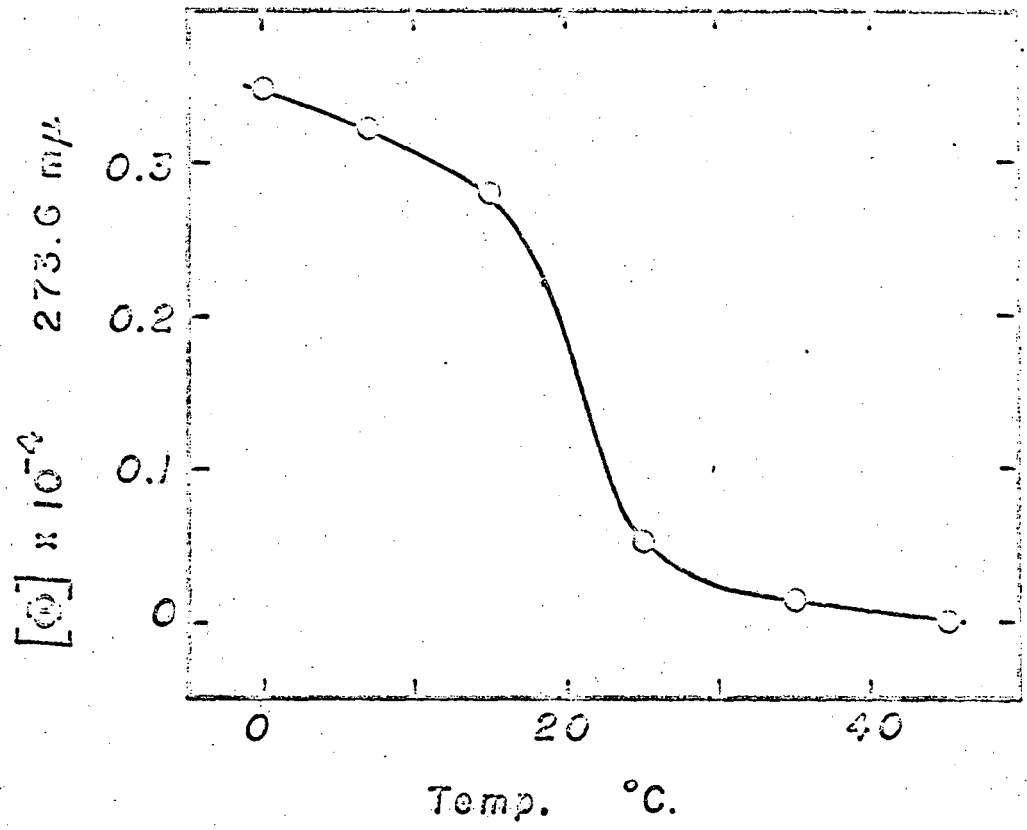


Figure 11. Melting curve for the omega-mer at the crossover wavelength of the 45°C ORD spectrum.

This curve should measure the melting of double strand regions in the omega-mer.

$$\mathcal{Y}_1 = e_1^{-1/2} \sum_i^n u_{1i}\, \mathcal{S}_i \quad ; \quad \mathcal{Y}_2 = e_2^{-1/2} \sum_i^n u_{2i}\, \mathcal{S}_i$$

where the quantities in the above expression are defined in Appendix II. This particular pair of basis vectors is purely an arbitrary choice. One can rotate these basis vectors, in the plane which they define, to obtain a new set of basis vectors, $\mathcal{Y}_1'$ and $\mathcal{Y}_2'$, such that $\mathcal{Y}_1'$ lies along the direction of $\mathcal{S}_{45}'$. $\mathcal{S}_{45}'$ is defined to be the component of $\mathcal{S}_{45}$ which lies in the plane of $\mathcal{Y}_1$ and $\mathcal{Y}_2$, where, in general, $\mathcal{S}_T$ is the vector representation of a spectrum taken at T°C. The magnitudes of the components of the set of $\mathcal{S}_T$ along the two basis vectors, $\mathcal{Y}_1$ and $\mathcal{Y}_2$, will be given by

$$\mathcal{S}_T \cdot \mathcal{Y}_1 = e_1^{1/2}\, u_{1,T} \quad \text{and} \quad \mathcal{S}_T \cdot \mathcal{Y}_2 = e_2^{1/2}\, u_{2,T}$$

Under the rotation described above, these components will transform according to the relation,

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \cdot \begin{bmatrix} \mathcal{S}_T \cdot \mathcal{Y}_1 \\ \mathcal{S}_T \cdot \mathcal{Y}_2 \end{bmatrix} = \begin{bmatrix} \mathcal{S}_T \cdot \mathcal{Y}_1' \\ \mathcal{S}_T \cdot \mathcal{Y}_2' \end{bmatrix} ,$$

where cos θ and sin θ can be calculated from the equation,

$$0 = \underset{\sim}{S}_{45} \cdot \underset{\sim}{V}_2' = (\cos\theta)\, e_2^{1/2}\, u_{2,45} - (\sin\theta)\, e_1^{1/2}\, u_{1,45}$$

The basis vector $\underset{\sim}{V}_1'$ now defines the shape of the spectrum taken at 45°C. Thus, the magnitudes of the set of $\underset{\sim}{S}_T$ which lie along $\underset{\sim}{V}_2'$ will be the points in the melting curve of the double strand conformation.

The melting curve obtained in the above manner is shown in Figure 12. One should note that, while the basis vectors have been chosen to be orthonormal, the vectors representing the single and double strand conformations will not be orthonormal. Thus, while the shape of the melting curve in Figure 12 has physical significance its magnitude is arbitrary. For the same reason, the magnitudes of the components of the set of $\underset{\sim}{S}_T$ along $\underset{\sim}{V}_1'$ will not give the melting curve of the single strand conformation, but rather some unknown linear combination of the melting curves of both conformations. Also, the shape of the ORD spectrum of the double strand conformation will not be defined by $\underset{\sim}{V}_2'$. In order to determine the shape of the ORD spectrum of the double strand conformation, it is necessary to know the temperature dependence as well as the shape of the ORD of the single strand conformation. The data, from the ORD spectra of the dinucleoside phosphates, necessary to calculate this information is available.[42] However, until one obtains a more complete

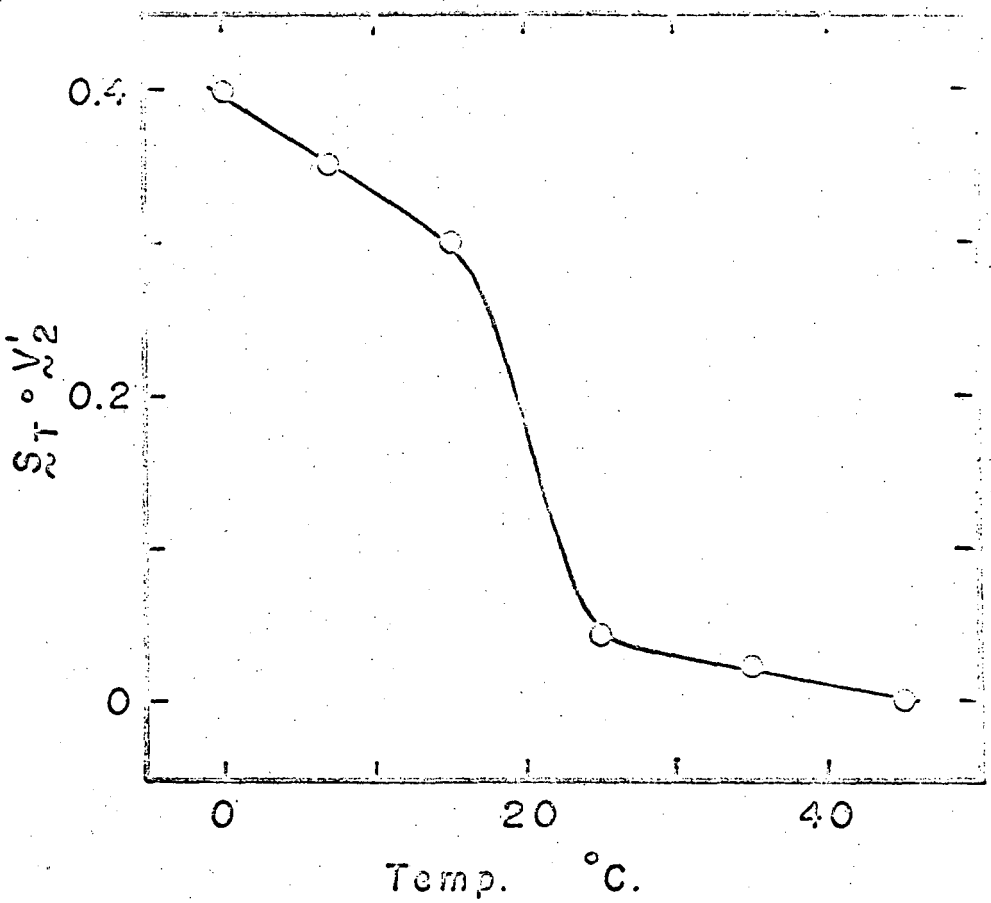


Figure 12. Melting curve for the second basis component of the omega-mer ORD spectra.

This curve should measure the melting of double strand regions in the omega-mer.

set of ORD data for the omega-mer, it is hardly worthwhile to perform this calculation. One could also use the data in the melting curve in Figure 12 to calculate thermodynamic quantities for the transition from the double to the single strand conformation. However, since so few points were obtained, none of which are at the midpoint of the curve, the values which one could obtain would not be meaningful.

If further studies were to indicate that the third component, which was ignored in the above treatment, was in fact significant, one might attribute such a small component to the ORD of the random coil conformation. One could verify such an assignment in the following manner, and at the same time reduce the set of spectra to a two component system. If one were to hydrolyze a sample of omega-mer and measure the temperature dependence of the ORD of its constituent monomers, the spectra obtained could be subtracted from those of the intact polymer. If the assignment were correct, the data would behave as a two component system in the eigenvalue analysis.

IV. PANCREATIC RNASE FINGERPRINTS OF THREE UNIQUE OLIGOMERS

1. Hydrolysis Conditions

The tally of fragments produced by treating a polynucleotide with a nuclease is commonly referred to as the fingerprint of the polynucleotide. One would like to use a nuclease which cleaves the RNA at specific positions to produce a well-defined set of possible fragments. This situation facilitates the identification of the hydrolysis fragments and allows the hydrolysis conditions to be reproduced accurately. Unfortunately, only two such ribonucleases have been well-documented. One of these is T-1 ribonuclease which has been found to cleave the phosphodiester linkage following a guanine residue.[43] Since the oligonucleotides being studied are the products of T-1 ribonuclease action, the only remaining nuclease suitable for fingerprint work is pancreatic ribonuclease.

The specificity of pancreatic RNase for the cleavage of phosphodiester bonds following pyrimidine residues has been demonstrated by a number of workers.[44,45,46] The result of the cleavage is an oligonucleotide ending in a terminal 2',3' cyclic phosphate which is converted by a second slower reaction, also mediated by the enzyme, to a terminal 3' phosphate.

The hydrolysis conditions initially chosen to give cleavage after all pyrimidine residues and conversion to the

3' phosphate were taken from a paper by Rushizky and Sober.[47] A sample containing 0.5 - 1.5 mg of the oligomer to be hydrolyzed was dissolved in 0.2 ml distilled water. To this solution was added 0.005 ml 1.0 M phosphate buffer, pH 7.9 and a quantity of a 1% solution of pancreatic ribonuclease (Worthington) sufficient to give an enzyme:substrate ratio of 1 mg:10 mg. The hydrolysis was allowed to proceed at 37°C for 20 hours.

The unsuitability of the above reaction conditions did not become apparent until a satisfactory separation technique was devised. The problem of the separation of the digestion products will be discussed in Chapter IV.2. It was eventually found that the hydrolysis conditions being used resulted not only in cleavage after pyrimidines, but also after adenosine. This activity has been previously reported[48] for pancreatic ribonuclease with large amounts of enzyme.

Table IV summarizes the types of cleavage observed. No cleavage was observed after an adenosine residue which was followed by a pyrimidine and the greatest amount of cleavage was observed with $(Ap)_3Cp$ which was principally degraded to ApAp and ApCp. It was not determined whether the terminal phosphate of the terminal adenosine was in the open or cyclic form.

As a result of the above observations, a milder set of hydrolysis conditions was introduced. The enzyme:substrate

TABLE IV

CLEAVAGES RESULTING IN A TERMINAL Ap OBSERVED WITH PANCREATIC RIBONUCLEASE

Hydrolysis Conditions:

2.5 - 7.5 mg RNA/ml in 0.025 M PO_4, pH 7.9

1:10 Enzyme:Substrate ratio (by weight)

20 hour incubation at 37°C

Initial Oligomer	Products	% Hydrolysis Observed
ApCp ApUp	None	0
ApApCp	Ap, ApCp	10 - 15%
ApApUp	Ap, ApUp	
ApApApCp	Ap, ApApCp	8 - 10%
	ApAp, ApCp	32 - 40%

ratio was reduced to 1 mg:20 mg and the reaction time was reduced to 3.5 hours. Under these conditions only 1 - 2% hydrolysis of $(Ap)_3Cp$ was observed and less than 1% of the uridylic acid remained as the cyclic diester. The results discussed in later sections were all obtained with the revised hydrolysis conditions.

2. Separation Techniques

In order to perform the fingerprinting described in the preceding section, it is necessary to have a reliable and relatively easy method for separating the digestion products produced by the action of pancreatic ribonuclease on a T-1 ribonuclease fragment. The two-dimensional, paper electrophoresis-chromatography map published by Rushisky and Sober[49] at first appeared to be ideal for this task.

A sheet of Whatman 3 MM paper (48 cm × 80 cm) was soaked with 0.1 M formic acid (pH 2.3 - 2.4) and allowed to equilibrate in an electrophoresis tank, also containing 0.1 M formic acid, for one hour. A circular area roughly 3 cm in diameter was blotted with 3 MM paper, and the hydrolysate was spotted on this area in a volume of 0.01 ml. A potential of 3000 volts was applied across the length (80 cm) of the paper for 2.5 hours. The paper was removed from the tank and dried. Descending chromatography was performed at ninety degrees to the direction of electrophoresis. The chromatogram was developed for 30 hours with a solvent consisting of 60% tert butanol: 40% 0.05 M NH_4OOCH,

pH 3.8. The positions of compounds on the paper were located by illuminating the chromatogram with UV light. Fig. 13 shows a trace of the map produced by this procedure. Although the map was capable of resolving the expected fragments, there appeared to be small amounts of unidentified fragments contained in the digest which were not resolved from the main spots. In particular, the spot which should have contained Cp was so badly contaminated that it was impossible to be sure that Cp was actually present. There was also a small spot in the lower left corner of the map which could not be identified at all. As mentioned in the preceding section, most of the above problems were eventually traced to the use of improper hydrolysis conditions. However, this was not possible until an improved separation procedure was discovered.

It is possible that another set of chromatography and electrophoresis conditions could have been found which would have given complete resolution. However, there were two other difficulties presented by paper chromatography-electrophoresis systems in general. In order to determine quantitatively the amount of a fragment present on the map by its UV absorbance, it was first necessary to elute the material from the paper. This is a tedious process; and, as shown by Fig. 14, 5' guanidylic acid could be recovered from the paper in yields of only 90%. A more serious problem was the occurrence of impurities, some of which were

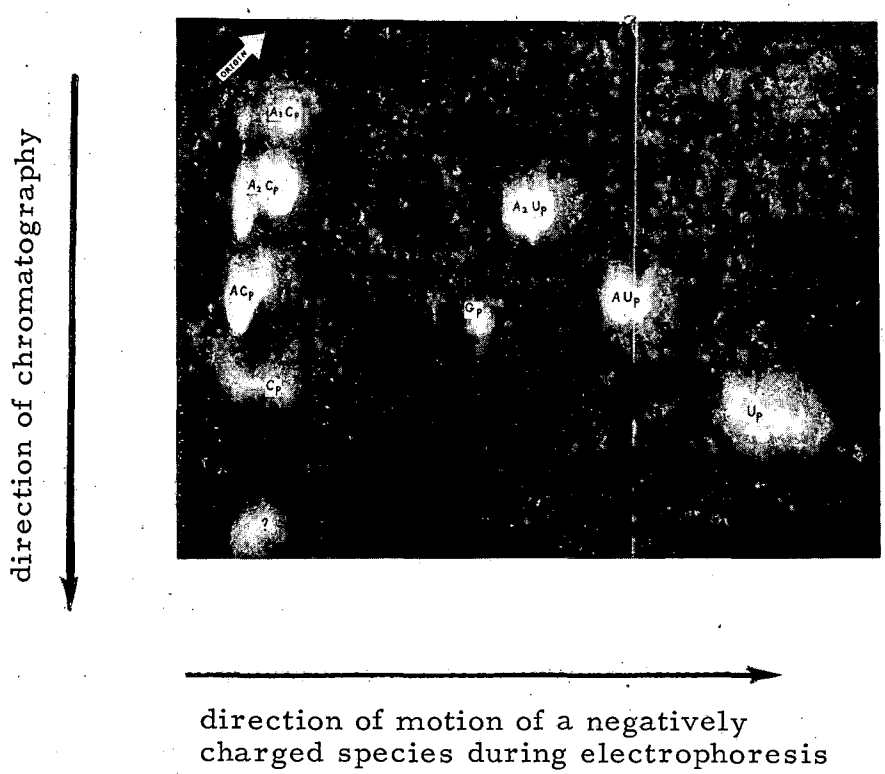


Figure 13. A paper electrophoresis-chromatography map of the pancreatic ribonuclease digestion fragments of the omega-mer.

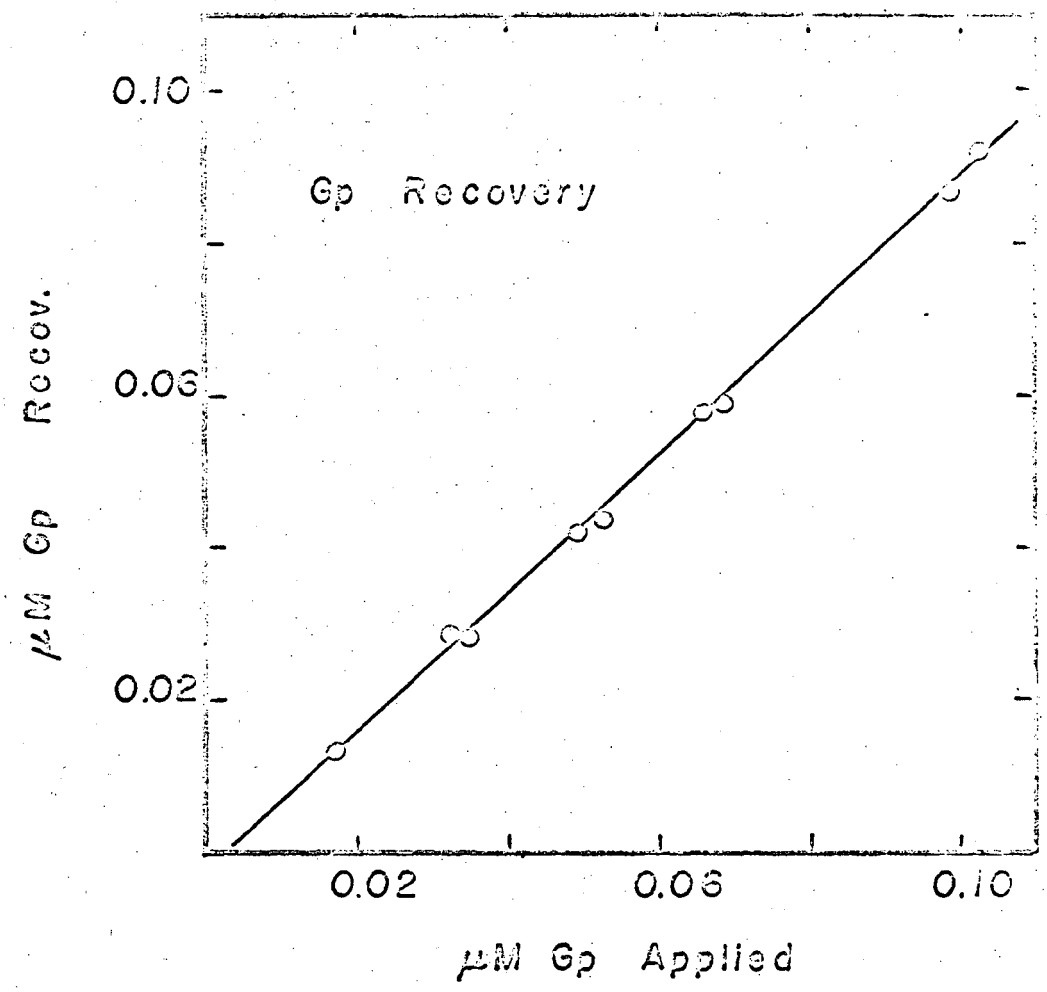


Figure 14. Recovery of Gp from paper electrophoresis-chromatography maps.

present in the paper originally and were difficult to remove, and others which were introduced during electrophoresis. These impurities gave UV absorption base lines which were erratic and therefore very difficult to correct. In order to circumvent the above difficulties a column chromatographic procedure using DEAE Sephadex was devised.

A DEAE Sephadex column (0.4 × 29 cm) was loaded with 10 - 50 mμ absorbance units of hydrolystate. The column was then developed with 30 ml 0.001 M HCl at a flow rate of 15 ml/hr before initiating a pH gradient. Although no material is eluted from the column, this step is a necessary part of the elution procedure. If the gradient is begun immediately, Cp and ApCp may elute together. A linear pH gradient was run at a flow rate of 1.7 ml/hr from 0.001 M HCl to 0.026 M HCl, 0.01 M KCl in a volume of 250 ml. At approximately the midpoint of the pH gradient a salt gradient was superimposed upon it by adjusting the reservoir chamber of the gradient apparatus to a concentration of 0.1 M in KCl. For maximum resolution this gradient change should be introduced after Up has been eluted from the column.

The identities of oligomers eluted from the column were assigned from their elution positions and a knowledge of the specificities of T-1 and pancreatic ribonuclease. These assignments were checked by comparing the UV absorption spectra of the oligomers eluted from the column

with those of known oligomers and, where necessary, by rechromatography with known oligomers. In all cases the results led to an unambiguous identification. The elution profiles shown in Figs. 15, 16, and 17 show the elution positions of the fragments obtained by pancreatic ribonuclease digestion of the psi-1-mer, psi-2-mer, and omega-mer.

The quantitative determination of oligomers which elute together, such as A_3Cp and Gp, can be done either by analyzing the UV absorption spectrum of the mixed oligomers or by rechromatography on DEAE-Sephadex at neutral pH with a salt gradient, as shown in Fig. 18. Between 97% and 98% of the absorbance units placed on the column were recovered in the eluent. The slight asymmetry observed in the elution profiles of the oligomers which eluted before the gradient change is apparently an artifact of the elution procedure rather than evidence of heterogeneity. When these peaks were rechromatographed at neutral pH they each gave an elution profile which contained a single symmetric peak.

This separation technique is actually more versatile than is indicated by the above figures. Fig. 19 shows the elution pattern obtained from a pancreatic ribonuclease digest of mixed T-1 ribonuclease oligomers of TMV-RNA. One sees that it is also possible to resolve oligomers of the form $(Ap)_nGp$.

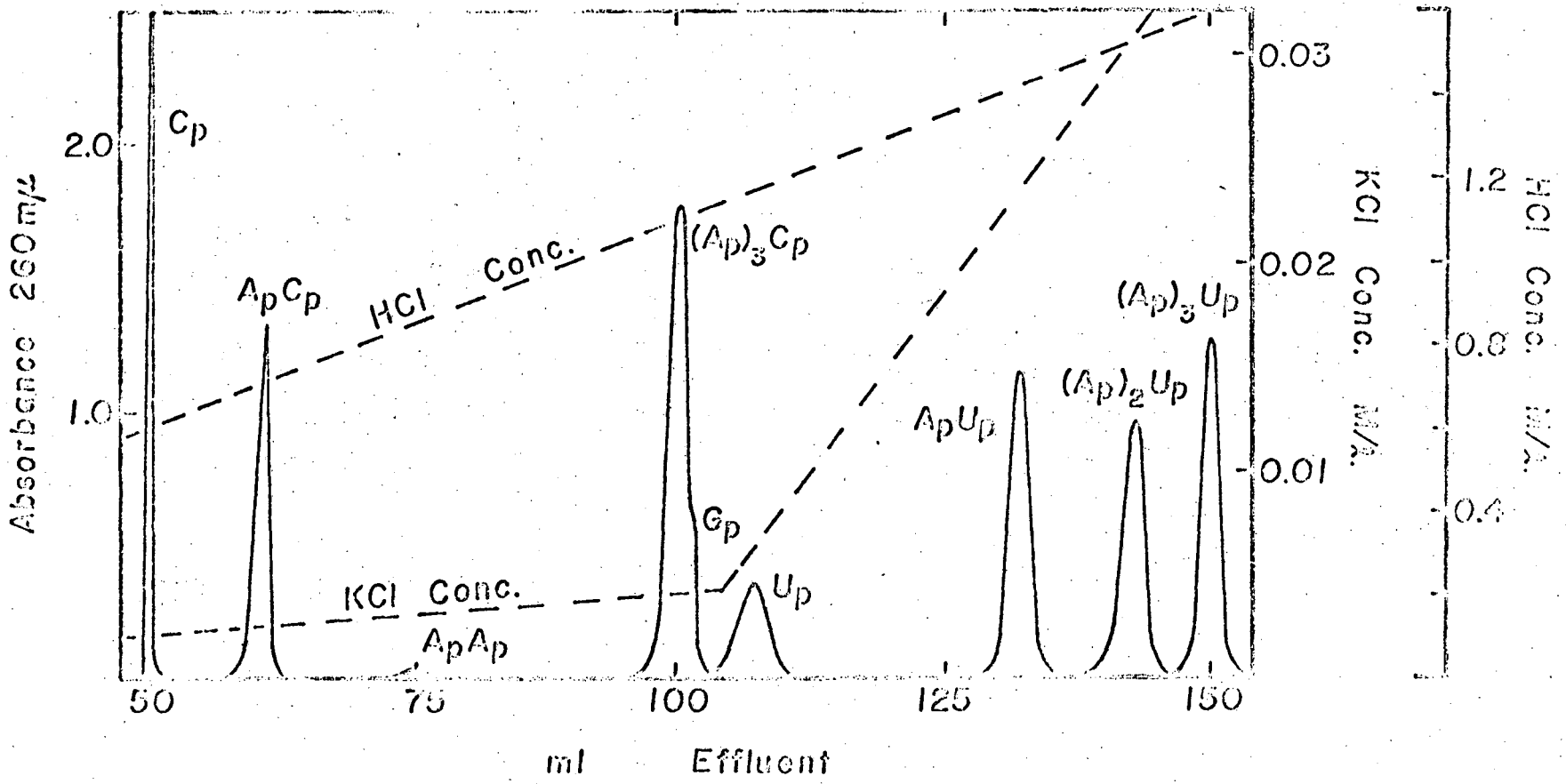


Figure 15. Separation of the pancreatic ribonuclease digestion fragments of the psi-1-mer.
Resin: DEAE-Sephadex, A-25
Bed Volume: 4 ml
Sample Load: 20 $A^{cm^2}_{260}$
Flowrate: 1.7 ml/hr

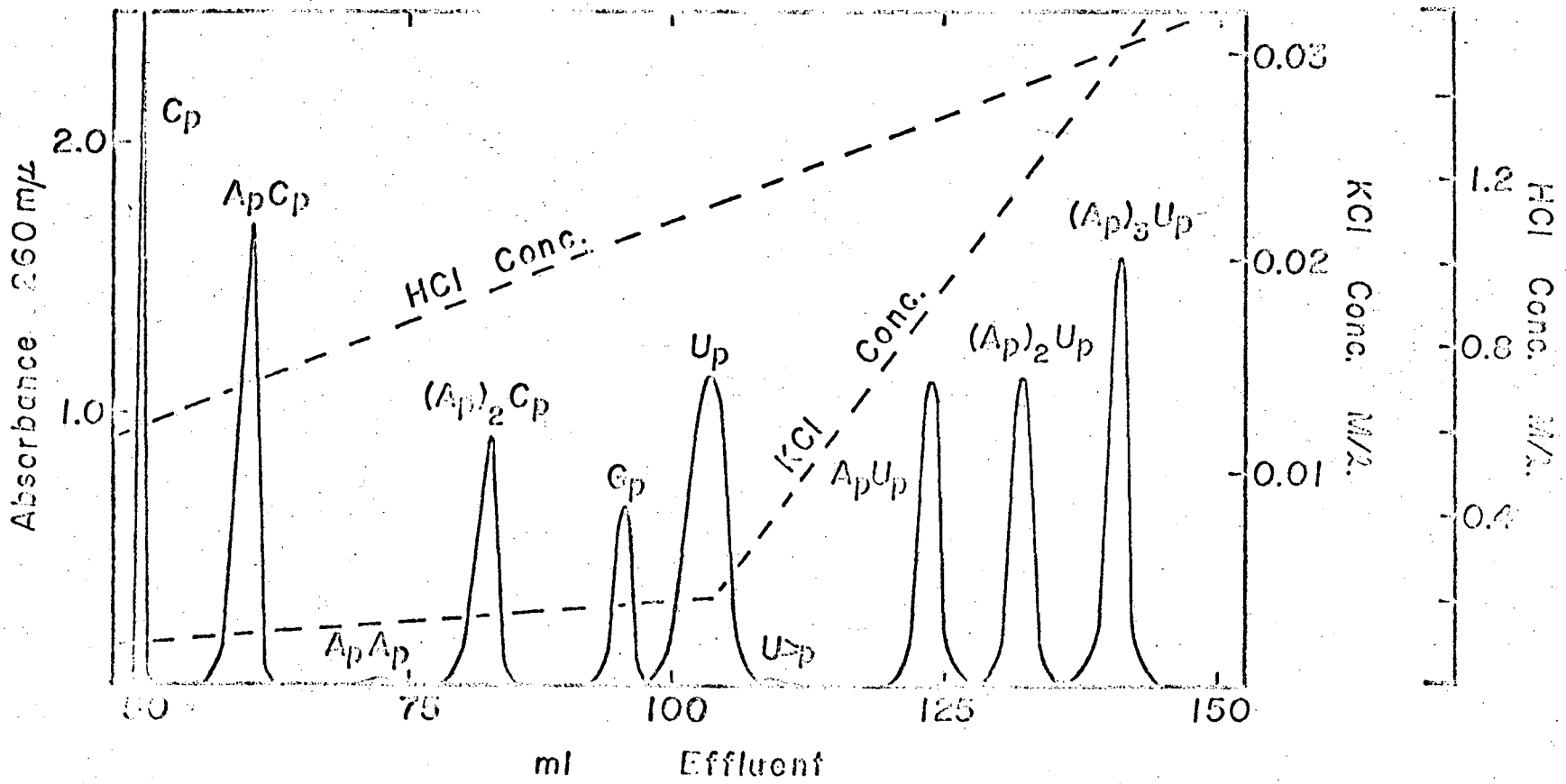


Figure 16. Separation of the pancreatic ribonuclease digestion fragments of the psi-2-mer.
Resin: DEAE-Sephadex, A-25
Bed Volume: 4 ml
Sample Load: 20 $A_{260}^{cm^2}$
Flowrate: 1.7 ml/hr

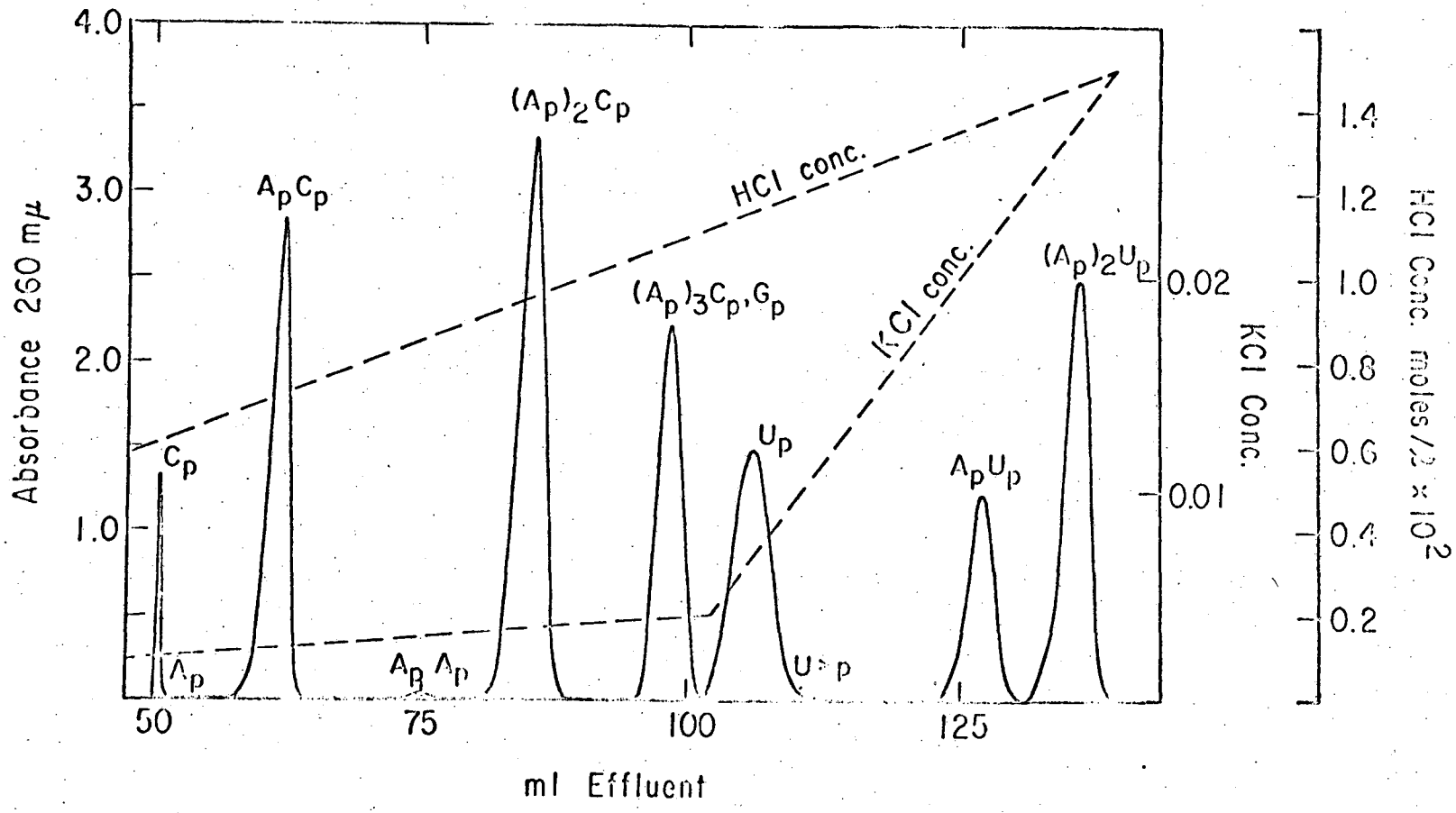


Figure 17. Separation of the pancreatic ribonuclease digestion fragments of the omega-mer.

Resin: DEAE-Sephadex, A-25
Bed Volume: 4 ml
Sample Load: 50 $A_{260}^{cm^2}$
Flowrate: 1.7 ml/hr

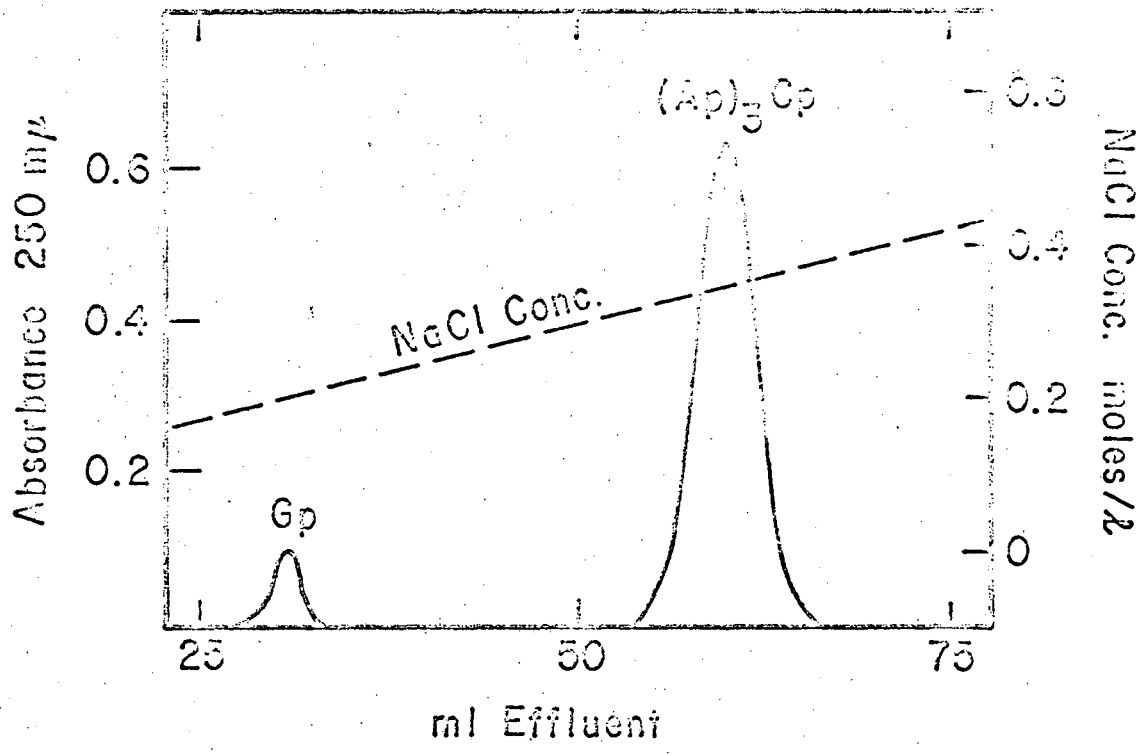


Figure 18. Rechromatography of the Gp, $(Ap)_3Cp$ peak from Figure 17.

Resin: DEAE-Sephadex, A-25
Bed Volume: 1 ml
Sample Load: 6 $A^{cm^2}_{260}$
Flowrate: 1.2 ml/hr

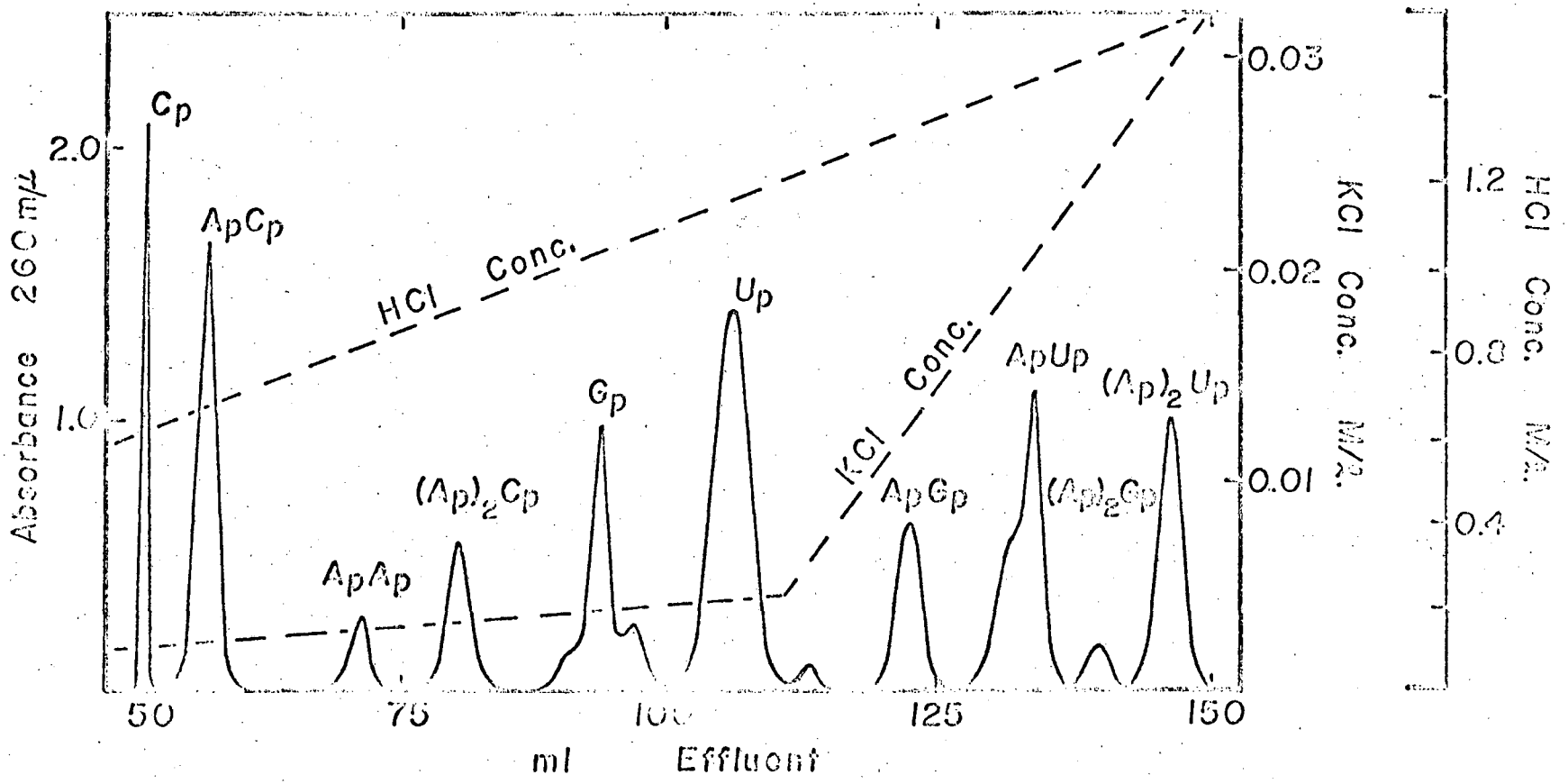


Figure 19. Separation of the pancreatic ribonuclease digestion fragments of mixed T-1 oligomers.

Resin: DEAE-Sephadex, A-25
Bed Volume: 4 ml
Sample Load: 1 mg
Flowrate: 2 ml/hr

3. Fingerprint Results

Pancreatic ribonuclease digests of the psi-1, psi-2, and omega-mers were fractionated by the column chromatography procedure of the preceding section. The portions of the eluent containing the digestion products were quantitatively collected, and their volumes were accurately measured. The UV-absorption spectrum of each fraction was measured and the number of 260 mμ absorbance units present in the fraction was computed. These values were converted to micro-molar values with the extinction coefficients shown in Table V. The three oligomers being investigated are the products of complete digestion with T-1 ribonuclease which causes cleavage after guanine[43] and should, therefore, each contain a single, terminal, guanine residue. Thus, the mole ratios of each oligomer isolated from the pancreatic ribonuclease digest, relative to guanadylic acid, should have integral values equal to the number of moles of each oligomer per mole of psi-1-mer, psi-2-mer, or omega-mer. Table VI shows the mole ratios obtained for two different digests of each of the three T-1 ribonuclease fragments. As mentioned in the preceding section, $(Ap)_3Cp$ and Gp were eluted as a single peak from the initial column and were separated by rechromatography at neutral pH. The results of this rechromatography were corrected to 100% yield (from 95%) for comparison with the values obtained from the original column. The small amounts of uridine

TABLE V

Spectral Properties of Oligonucleotides at pH 7.5

Oligonucleotide	Reference	Extinction coefficient at 260 mμ	A_{280}/A_{260}	A_{250}/A_{260}
	*	....	0.679±0.14	1.149±.017
Gp	Warshaw[51]	11.5	0.66	1.16
	*	....	0.944±.003	0.864±.003
Cp	Warshaw[51]	7.5	1.00	0.83
	*	....	0.426±.004	0.809±.009
ApCp	Warshaw[51]	21.0	0.44	0.80
	*	....	0.381±.007	0.823±.004
$(Ap)_2Cp$	Cantor[41]	30.6	0.423	0.827
	Stanley[50]	....	0.39	0.83

TABLE V (continued)

Oligonucleotide	Reference	Extinction coefficient at 260 mμ	A_{280}/A_{260}	A_{250}/A_{260}
$(Ap)_3Cp$	*	41.6 (calculated)	0.369±.013	0.830±.009
Up	*	....	0.369±.009	0.750±.003
	Warshaw[51]	9.9	0.38	0.73
ApUp	*	....	0.264±.005	0.770±.005
	Warshaw[51]	24.0	0.27	0.77
$(Ap)_2Up$	*	....	0.275±.004	0.800±.006
	Cantor[41]	33.9	0.326	0.813
	Stanley[50]	....	0.28	0.81
$(Ap)_3Up$	*	43.1 (calculated)	0.278±.004	0.816±.004

*Refers to values obtained in this work.

cyclic phosphate present are included in the uridine 3' phosphate tally. The values given for ApAp were estimated to ±20% from the area under the elution profile. The Cp eluted from the column was contaminated with small amounts of Ap. The amount of Cp present was calculated from UV absorption spectra using the 280/260 and 250/260 ratios given in Table V. Since the spectra of Cp and Ap are quite distinct, and since the amount of Ap present was small, this procedure gives very accurate values for Cp but rather poor values for Ap (about ±20%). Between 97% and 98% of the UV-absorbing material present in the pancreatic ribonuclease digests is accounted for by the oligomers shown in Table VI. Thus, it is reasonable to assume that no digestion fragments have been overlooked.

Table V also compares the values of the spectra ratios obtained for various oligomers with those reported by other workers. There is no evidence, either from the spectroscopic data or from rechromatography of the various oligomer fractions, which indicates the presence of odd bases in the hydrolysates.

The mole ratios given for the omega-mer show deviations from integral values of only 1% - 2% which is probably close to the limiting precision of the techniques employed. For the psi-mers, on the other hand, although the deviations between the two measurements are still on the order of 1% - 2%, the deviations from integral values are in some cases as high as 8%. These large deviations are probably

TABLE VI

Pancreatic Fingerprints of the Omega-mer and Psi-mers

digestion fragment	omega-mer uM fragment* per uM Gp	omega-mer closest integral value	psi-1-mer uM fragment* per uM Gp	psi-1-mer closest integral value	psi-2-mer uM fragment* per uM Gp	psi-2-mer closest integral value
Gp	1.00	1	1.00	1	1.00	1
Cp	1.05	1	3.86	4	2.08	2
ApCp	6.08	6	1.94	2	2.02	2
$(Ap)_2Cp$	6.04	6	not present	0	1.06	1
$(Ap)_3Cp$	1.96	2	0.92	1	not present	0
Up	12.00	12	2.04	2	5.30	5
ApUp	2.96	3	1.90	2	2.00	2
$(Ap)_2Up$	3.96	4	0.98	1	1.05	1
$(Ap)_3Up$	not present	0	0.93	1	1.04	1
Ap	0.10	0	0.04	0	0.05	0
ApAp	0.03	0	0.02	0	0.02	0
	$A_{35}U_{19}C_{15}G$		$A_{12}U_6C_7G$		$A_{11}U_9C_5G$	

*This figure is the average of the values obtained from two experiments.

due to oligonucleotide impurities since it would appear that even after rechromatography the psi-mer preparations are not as pure as those of the omega-mer. Compare Fig. 4 with Fig. 5. It is possible that these impurities are cleavage products of the psi-mers themselves, produced by hydrolysis at the acidic conditions at which the separation of the psi-mers is effected. However, despite the larger deviations observed for the psi-mers, the data presented are entirely satisfactory for the fingerprint determination. The results of the hydrolyses yield the pancreatic ribonuclease fingerprints shown in Table VI.

Thus the omega-mer has a chainlength of 70, which is in agreement with the sedimentation equilibrium measurement, and the psi-mers both have chainlengths of 26. The information gained from these fingerprints allows one to draw a number of conclusions about the biological function of certain regions of the viral RNA. These conclusions will be the subject of Chapter VI.

V. ATTEMPTS TO OBTAIN C-SPECIFIC CLEAVAGE

1. Carbodiimide Blocking Groups

Pancreatic ribonuclease digestion of the omega-mer and psi-mer produces such a large number of fragments that ordering them by any sort of partial digestion technique would be a very difficult task. A possible method of simplifying the problem would be to devise a way of achieving cleavage only after cytidine. This, would, in general, lead to half the number of fragments produced by treatment with pancreatic ribonuclease. A series of papers by Gilham and his co-workers[52,53] describes a technique for achieving specific cleavage after cytosine with pancreatic ribonuclease by the addition of a carbodiimide derivative to the uracil ring which then protects these sites from the enzyme.

The specificity of the action of Gilham's reagent (1 cyclohexl-3-(2-morpholinoethyl)-carbodiimide metho-p-toluenesulfonate, Aldrich Chem. Co.) on the mononucleotides was verified by the following experiment.

Four samples, each containing 10 mg of one of the four nucleotides Ap, Gp, Cp, Up, were dissolved in four 2 ml portions of distilled water. To each of the four solutions was added 200 mg of Gilham's reagent and enough 0.1 M NaOH to bring the pH to 9.2. The reaction was allowed to proceed for 20 hours at 25°C. Electrophoresis was performed on

0.05 ml aliquots of each of the samples using the techniques described in Section 2 of the preceding chapter. A field of 3000 volts was applied for 1.5 hours across 80 cm of paper wet with 0.05 M NH_4OOCH, pH 2.7. Untreated samples of the four nucleotides were run as markers on the same paper.

The reaction mixtures containing Ap and Cp showed only those spots corresponding to the reagent and to unreacted Ap and Cp. By contrast, the other two solutions contained no detectable Up or Gp. Instead, each of these solutions contained, in addition to the unreacted reagent, a compound which moved towards the negative electrode in electrophoresis. This is the behavior which one would predict for the products formed by the addition of the positively charged reagent to Up and Gp. Each of the two positively charged species was eluted in approximately 0.5 ml of distilled water, and an equal volume of concentrated NH_4OH was added to each sample. After standing for 2 hours at 25°C the two preparations were subjected to electrophoresis under the conditions given above. After treatment with NH_4OH each sample showed a single spot corresponding to Up and Gp. Both the addition and removal of the carbodiimide appeared to be at least 99% complete.

Although the results of the reaction of Gilham's reagent with the mononucleotides were very encouraging, attempts to achieve C-specific cleavage of oligonucleotides were not nearly so satisfactory. In order to insure complete

reaction of the carbodiimide with an oligonucleotide one wants to destroy any secondary or tertiary structure possessed by the oligomer. Accordingly, the reaction was carried out in 2 ml 7 M urea, 10^{-3} M EDTA containing 20 mg of oligomers of chainlength 7 - 10 obtained by T-1 ribonuclease hydrolysis of TMV-RNA. Two hundred mg of Gilham's reagent was added and the pH was adjusted to 9.2 with 0.1 M NaOH. The reaction was performed at 25°C for 20 hours.

The reaction was terminated by adjusting the mixture to pH 7.2 with 0.1 M HCl. The urea and unreacted reagent were removed by successive dialysis for 4 hour periods against 4 liters 0.1 M NaCl, 4 liters distilled water, and 4 liters 10^{-3} M PO_4, pH 7.2. After the addition of 0.04 ml 1.0 M PO_4, pH 7.2 and 0.1 ml of a 1% solution of pancreatic ribonuclease the sample was incubated for 20 hours at 25°C.

Before removing the carbodiimide from the uracil, it is first necessary to inactivate the ribonuclease. The enzyme was first denatured by allowing the sample to stand for 20 hours at 25°C after the addition of one volume (approximately 3 ml) dimethylformamide and 0.1 ml β-mercaptoethanol. During the denaturation process it is important that oxygen be excluded. This was accomplished by flushing the container with nitrogen before sealing it. The sample was evacuated to dryness and dissolved in 2 ml of distilled water which had been placed in an evacuated chamber to remove dissolved oxygen. The sample was adjusted

to pH 7.0 with 0.1 M NaOH, 0.1 ml of a 1% solution of pronase (Calbiochem, B grade) was added, and after it was flushed with nitrogen the container was again sealed. The pronase digestion of the ribonuclease was allowed to proceed for 24 hours at 25°C. The sample was finally lyophilyzed to dryness and taken up in 0.2 ml distilled water.

A 0.05 ml aliquot of the above solution was mapped in the paper electrophoresis-chromatography system described in Chapter IV.2. In this preliminary work no attempt was made to identify all of the fragments present in the map. The hydrolysates were principally examined for the presence of Up. If one succeeded in blocking the action of pancreatic ribonuclease adjacent to uridine residues, the only remaining sites susceptible to cleavage by this enzyme would occur on the 3' linked side of a cytidine residue. Thus, all of the digestion fragments resulting from pancreatic ribonuclease treatment would be of the form $(A_n,U_m)Cp$, except for terminal fragments of the form $(A_n,U_m)Gp$. It would be impossible to obtain Up as a digestion fragment. Therefore, the presence of Up in a hydrolysate would indicate that cleavage was occurring adjacent to uridine residues. The results obtained with the carbodiimide blocking group were not reproducible. Fig. 20 illustrates a trial hydrolysis which contained a large amount of uridylic acid while the result of the trial hydrolysis shown in Fig. 21 indicates that the blocking

origin

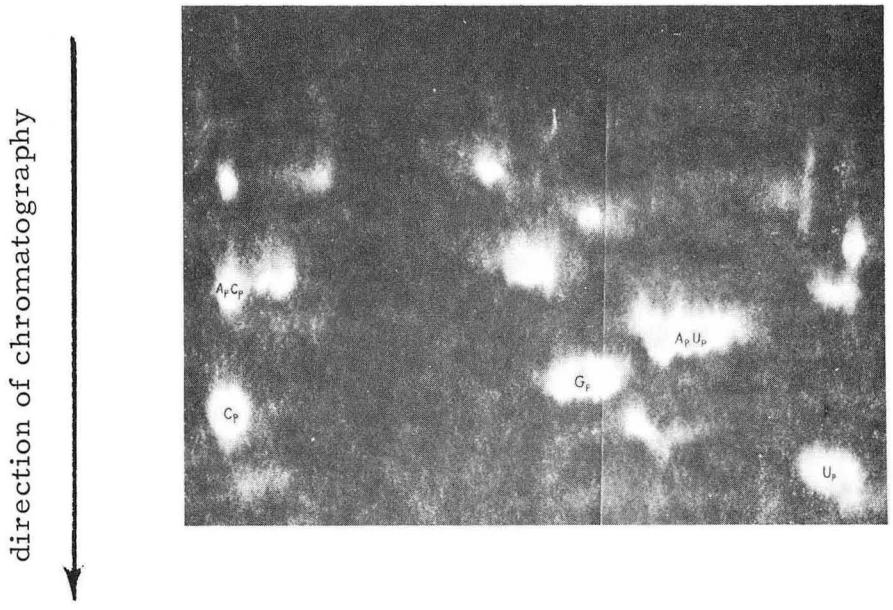


Figure 20. The result of pancreatic ribonuclease hydrolysis with the Gilham blocking reagent.

(Note the presence of Up in the digest.)

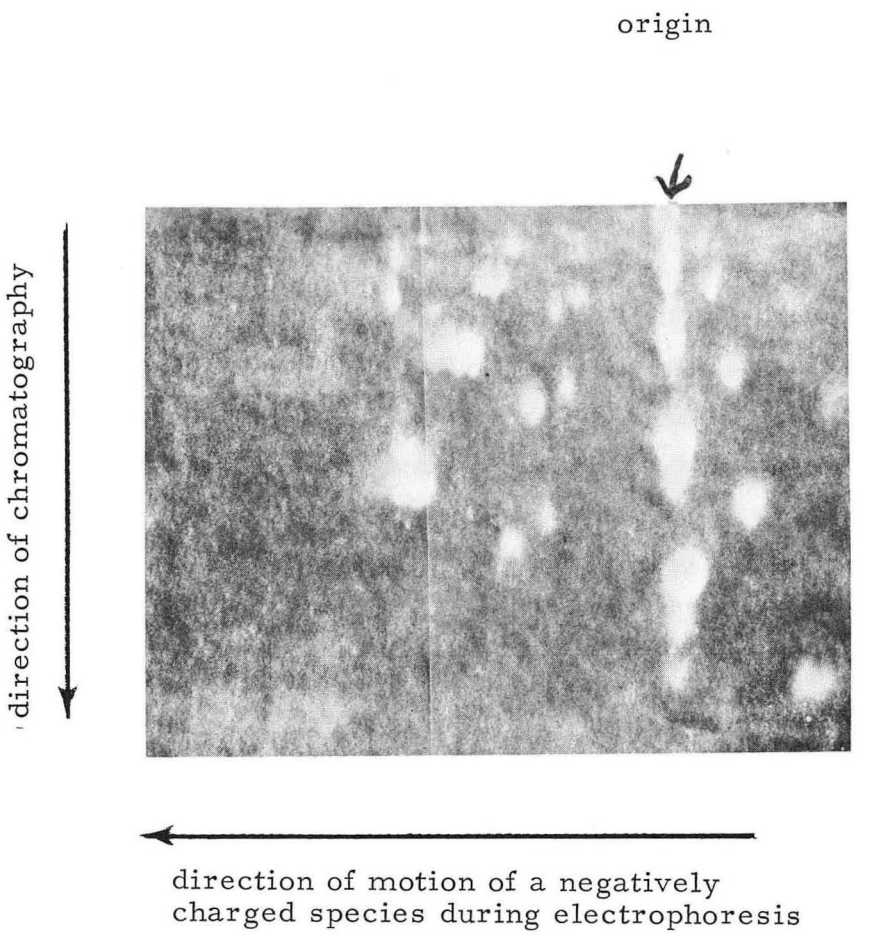


Figure 21. The result of pancreatic ribonuclease hydrolysis with the Gilham blocking reagent.

(Note the presence of positively charged fragments in the digest.)

group was not completely removed as evidenced by the presence of the positively charged components to the right of the origin.

In a more recent paper Gilham[54] also reports the presence of Up in his digests, but only in trace amounts. He reports no difficulty in removing the blocking group. If the cleavage after Up is due to incomplete reaction of the blocking group, one might solve the problem by running the reaction at a higher temperature. Mandeles[10] has reported success in removing the ribonuclease with bentonite which would considerably simplify the procedure given here. However, he also has experienced difficulty in the removal of the reagent with NH_4OH. Even the rather poor results presented here seem to indicate that Gilham's technique is a satisfactory method of preparing oligomers containing uridine with a terminal Cp. However, it became apparent that considerable time and effort would be required to develop a satisfactory analytical technique. For this reason, work on the carbodiimide reagent was abandoned in favor of the alternative to be described in Chapter V.2.

2. An Acid-Soluble Nuclease from E. Coli

At present only two of the ribonucleases have been clearly demonstrated to possess a marked specificity for cleavage of the phosphodiester bond adjacent to specific

bases. However, in 1965 Anderson and Carter[55] reported the isolation of an acid-soluble nuclease preparation, from ribosomes of E. Coli, whose rate of hydrolysis was much greater for poly C than that for either poly A or poly U. The same paper described a similar preparation which was obtained from the salt fractionated alkaline phosphatase (BAPSF) sold by the Worthington Biochem. Corp. Although the nuclease obtained from the latter source was capable of hydrolyzing poly C at a rapid rate, it showed no activity at all for poly A and poly U. Thus, it seemed worthwhile to investigate this nuclease preparation as a means of obtaining specific cleavage after cytidine.

The procedure used in isolating the nuclease was the same as that described by Anderson and Carter.[55] One vial of Worthington alkaline phosphatase (BAPSF) containing 10 mg of protein was dissolved in 5 ml distilled water at 0°C. One ml of 1.0 M $HClO_4$, also at 0°C, was added to precipitate the phosphatase. The precipitated protein was removed by centrifugation at 12,000 x g. for 5 minutes at 4°C, and the supernatant was adjusted to pH 7.5 - 8.0 with 1.0 M KOH. In order to insure maximum precipitation of the insoluble $KClO_4$, the neutralized solution was held at 0°C for 10 minutes before being centrifuged at 12,000 x g. for 10 minutes. The resulting nuclease preparation, designated BAPSF-Ac by Anderson and Carter, was stored frozen at -20°C.

The enzyme was assayed by following the release of acid-soluble products from polycytidylic acid. The assay mixture contained 3 ml of 0.067 mg/ml polycytidylic acid in 0.04 M Tris-Cl, pH 8.0, and 0.05 ml - 0.10 ml of BAPSF-Ac. After being incubated for 10 minutes at 37°C, the reaction was terminated by the addition of 1 ml 1.0 M $HClO_4$ at 0°C. The assay samples were kept at 0°C for 10 minutes to insure complete precipitation of the polynucleotide, which was then removed by centrifugation at 12,000 x g. for 10 minutes. The absorption of the supernatants was measured at 280 mμ. One unit of nuclease activity caused an increase of 0.1 in the 280 mμ absorbance under the above assay conditions. The specific activity was defined as the number of units of activity per mg of protein as determined by the Lowry[56] method.

The BAPSF-Ac fractions were found to contain about 0.1 mg/ml of protein, which is in agreement with the value of 0.092 mg/ml reported by Anderson and Carter. However, the specific activities obtained were on the order of 200 units/mg which was a factor of ten less than the value given by the above authors. It is possible that the difference in activities is due to variations in the amount of contaminating nuclease present in the Worthington phosphatase, although preparations made by this author from two different lots of BAPSF (6167 and 6GA) showed only small variations in activity. No attempt was made to verify the absence of anion and cation requirements

reported by Anderson and Carter for this enzyme.

The products of poly C hydrolysis were found to terminate in a 2', 3' cyclic phosphate. When the above assay was performed in a titrimeter in the absence of Tris-Cl buffer, no hydrogen ion was evolved over a period of 4 hours although the poly C was completely degraded to acid-soluble products. If the reaction which opens the cyclic phosphates proceeds at all, it does so at an immeasurably slow rate.

In an experiment designed to test the nucleotide specificity of the BAPSF-Ac fraction, four 0.05 ml portions of enzyme, each containing one unit of activity, were combined with 0.1 mg portions of poly A, poly U, and poly C, each in a volume of 0.05 ml of 0.04 M Tris-Cl buffer, pH 8.0. The reaction mixtures were incubated for 12 hours at 37°C and then spotted along with untreated samples of the three homopolymers on a 46 × 58 cm sheet of Whatman 3 MM paper. Descending chromatography was performed for 20 hours in a solvent system composed of 50% 1-propanol, 25% water, and 25% concentrated ammonium hydroxide. Although the amount of enzyme used in the reaction was capable of converting all of the poly C into acid-soluble products in about 2 hours, Fig. 22 shows that appreciable amounts of dimer, trimer, and tetramer remain in the digest even after 12 hours. One-half to two-thirds of the nucleotides were judged to be present as cytidine 2', 3' cyclic phosphate with the remaining material fairly

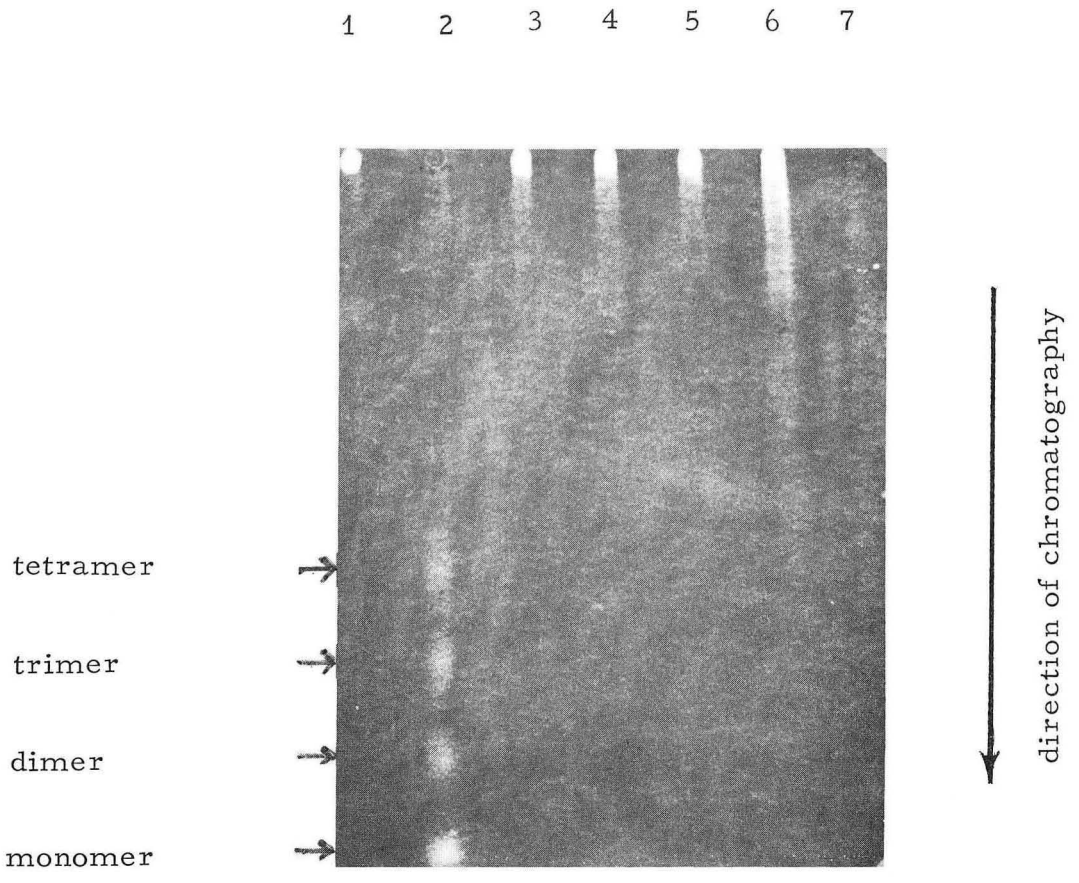


Figure 22. Chromatography of BAPSF-Ac digests of homopolymers.

1. poly C
2. poly C plus BAPSF-Ac
3. poly A
4. poly A plus BAPSF-Ac
5. poly U
6. poly U plus BAPSF-Ac
7. BAPSF-Ac

Developed for 20 hr with 2 : 1 : 1 : 1-propanol : water : ammonium hydroxide

equally distributed as dimer, trimer, and tetramer. It would seem that the BAPSF-Ac fraction is either incapable of hydrolyzing small homo-oligomers of cytidylic acid or does so at a greatly reduced rate compared to its action on poly-Cp. No degradation products were observed for either poly A or poly U. However, it was observed that the poly U sample treated with enzyme tailed from the origin slightly more than did the control. This effect may indicate that a small number of breaks occurred in the poly U, or it may only be an artifact since some tailing occurred in the control. An experiment similar to the one described above was used to test the specificity of BAPSF-Ac towards the four dinucleoside phosphates ApC, UpC, CpC, and GpC. Under reaction conditions identical to those described above no activity was observed for any of the four compounds. Thus, it appears that this enzyme preparation requires, at least, a trimer as substrate, and shows appreciable activity only for oligomers having chainlengths greater than four.

A 0.5 mg sample of T-1 ribonuclease fragments of TMV-RNA (of chainlengths of 7 - 10) in 0.05 ml 0.04 M Tris-Cl pH 8.0 was incubated with 0.3 ml of BAPSF-Ac containing five to six units of activity. The hydrolysis was performed at 37°C for a period of 20 hours. The hydrolysate was mapped with the paper electrophoresis-chromatography system described in Chapter IV.2. Fig. 23 shows the pattern of fragments obtained from the

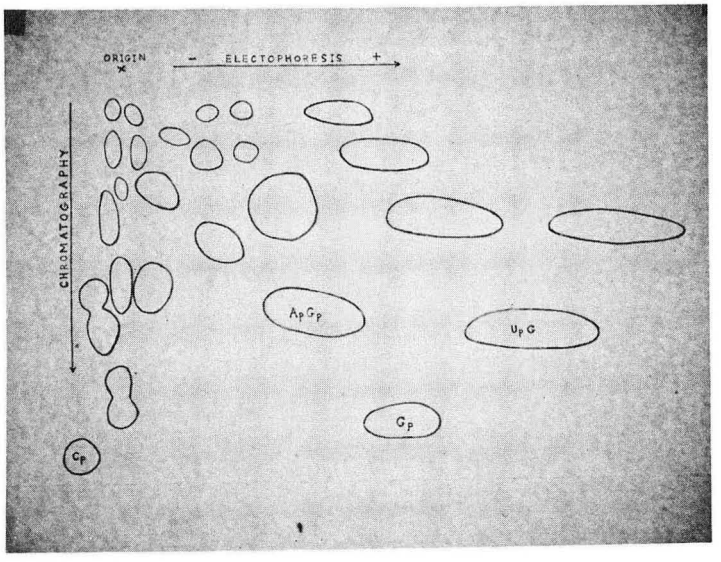


Figure 23. The hydrolysis of T-1 fragments with BAPSF-Ac.

hydrolysate. Only four of the fragments, Cp, Gp, ApGp, and UpGp, were definitely identified. The significant observation to be made is that Up, which would result from cleavage following a uridine residue, is absent from the digest. The portion of the map which should contain this nucleotide is completely blank.

Although a definitive test, in which all of the fragments of a BAPSF-Ac RNA hydrolysate are identified, has not been performed, it seems clear that for the enzyme concentrations used in these experiments no significant amount of hydrolysis is obtained after any nucleoside other than cytidine. Unfortunately, the cleavage after cytidine is not complete. It may be that trimers and tetramers can be hydrolyzed at higher levels of enzyme although cleavage after uridine may become a problem under these conditions. In any case, the dimers appear to be completely inert to the action of the enzyme. Thus, this nuclease preparation will probably not be as useful as T-1 and pancreatic ribonuclease; but it appears to offer a perfectly feasible way of obtaining C-specific cleavage.

3. Additional Sequence Studies on the Omega-mer

The results obtained with the BAPSF-Ac fraction described in the preceding section encouraged its use in two experiments with the omega-mer. The first of these was an attempt to obtain the fingerprint resulting from cleavage after cytidine. A 0.5 mg sample of omega-mer

was hydrolysed with 5 units of BAPSF-Ac in a volume of 0.3 ml 0.04 M Tris-Cl, pH 8.0. The reaction was carried out at 37°C for 20 hours. The hydrolysate was mapped in the paper electrophoresis-chromatography system which has been previously described. The pattern of spots obtained from the hydrolysate is shown in Fig. 24. The spots labeled 5, 6, and 7 were eluted from the paper in 0.05 ml of water and hydrolized with pancreatic ribonuclease. The solutions containing the fragments from the map were combined with 0.02 mg of enzyme, 0.005 ml of 1.0 M PO_4, pH 8.0, and digested for 20 hours at 37°C. The secondary hydrolysates were then mapped in the same system used above.

It was possible to identify some of the pancreatic ribonuclease fragments from their positions in the secondary maps. Spot number 5 was found to yield G and ApApUp while in the case of spot number 7 only ApApUp was observed in the secondary digest. The number of absorbance units contained in spot number 6 was so low that none of its pancreatic fragments could be identified. It is apparent that these results do not account for all of the fragments which one would expect fo find in the secondary digests. For example, during electrophoresis in the primary map, spot number 7 moved a distance greater than or equal to that which one would expect for the dimer ApUp. Yet, when the material contained in this spot was treated with pancreatic ribonuclease, the only detectable digestion product

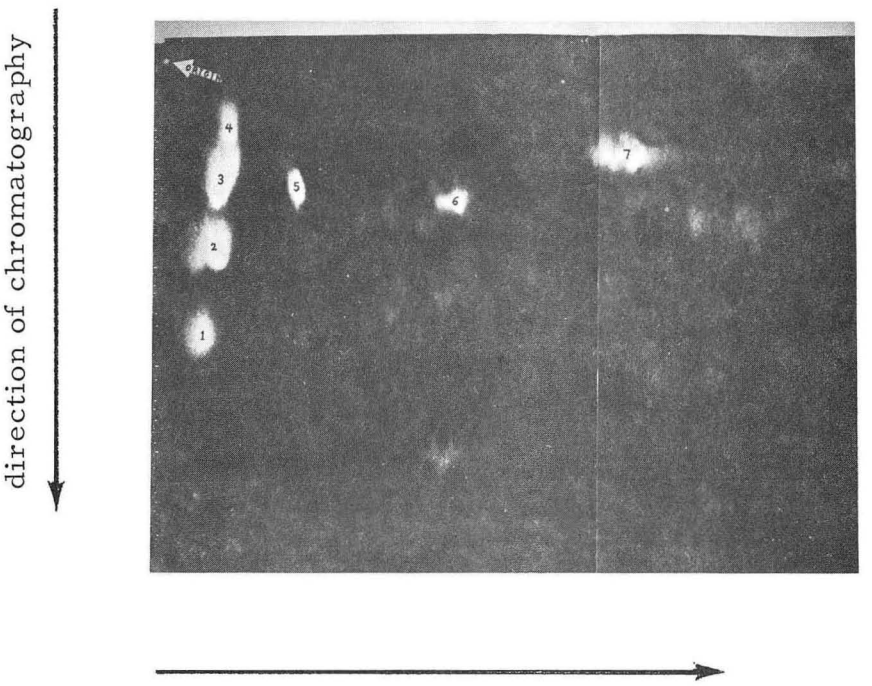


Figure 24. BAPSF-Ac digestion fragments of the omega-mer.

was ApApUp which, at pH 2.4, possesses an electrophoretic mobility much lower than that of ApUp. In order to account for the mobility of the oligomer present in spot number 7, it is necessary to include at least one additional uridine residue in its sequence. One concludes that small amounts of Up may have gone undetected in all of the secondary maps. The difficulty in detecting the uridine monophosphate fragments is that Up tends to form large diffuse spots which make it impossible to detect in small amounts.

The results of this experiment are such that it is impossible to determine whether any cleavage has occurred after uridine. In order to obtain clear results it will be necessary to perform this experiment with much larger amounts of material. It would also be desirable to use a column chromatographic separation scheme. The disadvantages of the paper electrophoresis-chromatography system were discussed earlier. It does appear from the above experiment that the terminal sequence of the omega-mer is ApApUpGp. However, it must be admitted that the position of spot number 5 is such that the sequences UpApApUpGp and ApApUpUpGp can not be ruled out.

Although the question of the specificity of the BAPSF-Ac preparation has not been definitely settled, it is clear that this enzyme preparation possesses a marked preference for cleavage after cytidine. Thus, it seemed to be a better choice than pancreatic ribonuclease for obtaining partial

hydrolysis of the omega-mer. It was hoped that, by performing hydrolyses for short periods of time, at low temperature, and in the presence of magnesium, the enzyme could be made to preferentially cleave a small number of diester linkages in the omega-mer. The reaction mixtures contained 0.3 mg omega-mer, 10 units of BAPSF-Ac, 10 μM Tris-Cl, pH 8.0, 40 μM KCl, and 4 μM $MgCl_2$ in a total volume of 1 ml. The hydrolysis was performed at 0°C.

A quantity of solid urea sufficient to bring the urea concentration to 7 M was added to each hydrolysate. It was then loaded on a 0.4 cm × 29 cm column packed with DEAE-Sephadex A-25 which had been previously equilibrated with 7 M urea. The column was eluted with a linear salt gradient running from zero to 0.8 M NaCl in 250 ml of 7 M urea. The flowrate was maintained at 2 ml/hr. Fig. 25 and 26 show the elution patterns obtained for hydrolysis times of 20 minutes and 7 hours.

The elution profiles indicate that essentially random hydrolysis is occurring. If the enzyme does possess a perference for cleavage at certain sites in the omega-mer, this preference is not large enough to allow one to obtain a significantly large yield of any one fragment. In theory, one such partial digest should contain all the information needed to reconstruct the entire sequence. In practice, unless one can obtain large yields of a few fragments, the process of sorting out a large number of

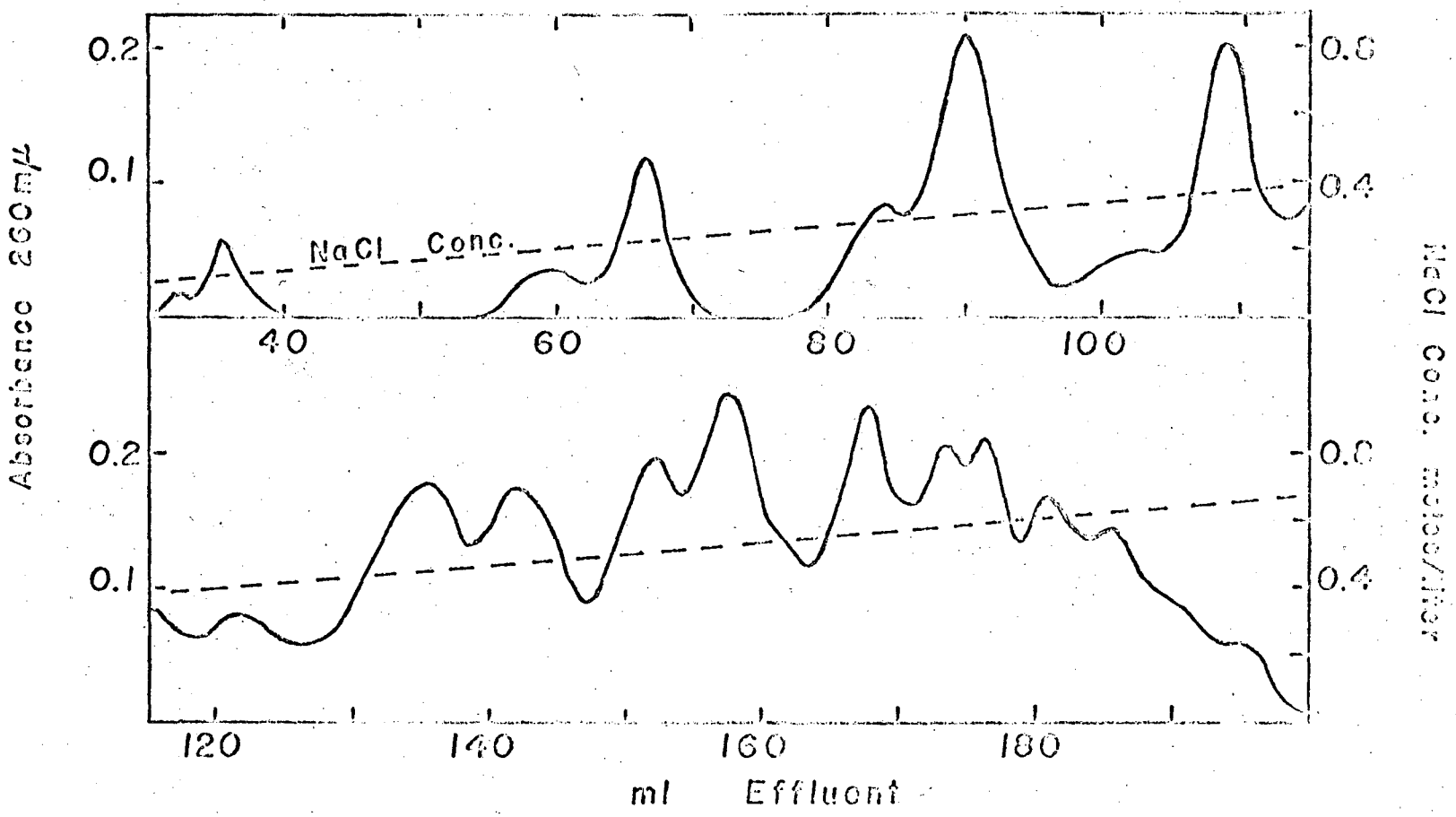


Figure 25. Partial hydrolysis of the omega-mer with BAPSF-Ac; 7 hr.

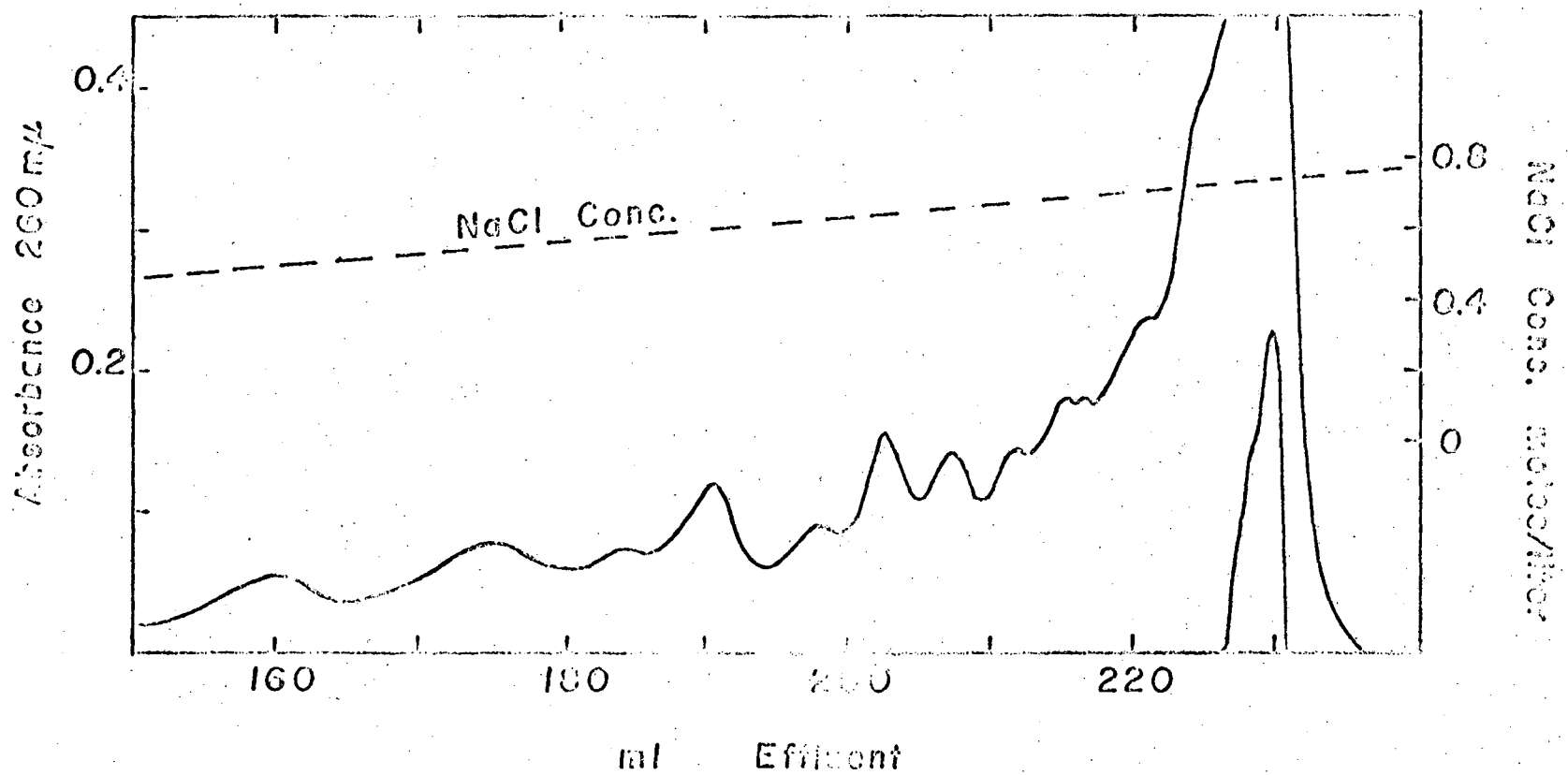


Figure 26. Partial hydrolysis of the omega-mer with BAPSF-Ac; 20 min.

fragments for further analysis is extremely difficult and requires large amounts of starting material. Thus, the method of partial hydrolysis used here does not seem particularly attractive. An alternate method is suggested in Chapter VI.3.

VI. DISCUSSION

1. The Location of Digestion Fragments in TMV-RNA and the Question of their Uniqueness

A series of experiments performed by Mandeles[12] have a very direct connection with the work reported here. It is desirable that these experiments should be discussed before proceeding with a discussion of the results of the fingerprinting experiments. It has been shown[57] that the action of sodium dodecylsulfate (SDS) on TMV is such that the coat protein is removed sequentially from one end of the virus rod. The RNA which is exposed always contains the end with the free 3' hydroxyl group.[58] The extent to which the coat protein is stripped from the virus can be controlled by the SDS concentration, the temperature, and the reaction time.[59] Thus, one can perform an experiment on a terminal segment of the TMV-RNA strand. This technique has been used by Kado and Knight[60] to determine the position of the gene responsible for the local lesion mutation of TMV.

In the work reported by Mandeles, three percent to fifty percent of the coat protein was stripped from a number of samples of TMV. The average amount of stripping was determined from the decrease in light scattering, the amount of 280 mμ absorbing material liberated from the virus, and the amount of RNA which became susceptible to hydrolysis by ribonuclease T-1. The partially stripped

virus was isolated and the exposed RNA was hydrolyzed with ribonuclease T-1. The unstripped portion of the virus was isolated and its RNA was extracted and hydrolyzed. The hydrolysates were chromatogramed on DEAE Sephadex under conditions similar to those described in Chapter II. The amounts of psi-mers and omega-mer present were determined for each hydrolysate. A plot of the amount of omega-mer or psi-mer released by T-1 ribonuclease hydrolysis as a function of the percent of protein stripped can be used to determine the approximate location of the oligomer in the RNA strand. Such a plot should have a sigmoid shape; and the percentage of protein stripped at the inflection point of the curve should give the position of the oligomer in the viral RNA.

Mandeles has placed the psi-1-mer only 360 nucleotides from the 3' end of the virus and the psi-2-mer roughly at the center of the RNA strand. The omega-mer was determined to be no more than 180 nucleotides from the 3' terminus. This figure is only an upper estimate since the smallest amount of stripping obtained was 3%, at which point almost all of the omega-mer was exposed. If one uses the curve in Mandeles' paper showing the release of psi-1-mer to make a guess at the width of the distribution curve for the stripping reaction, one concludes that the omega-mer is probably located only 80-100 nucleotides from the end of the RNA. Since this is an estimate of the position of the 5' end of the

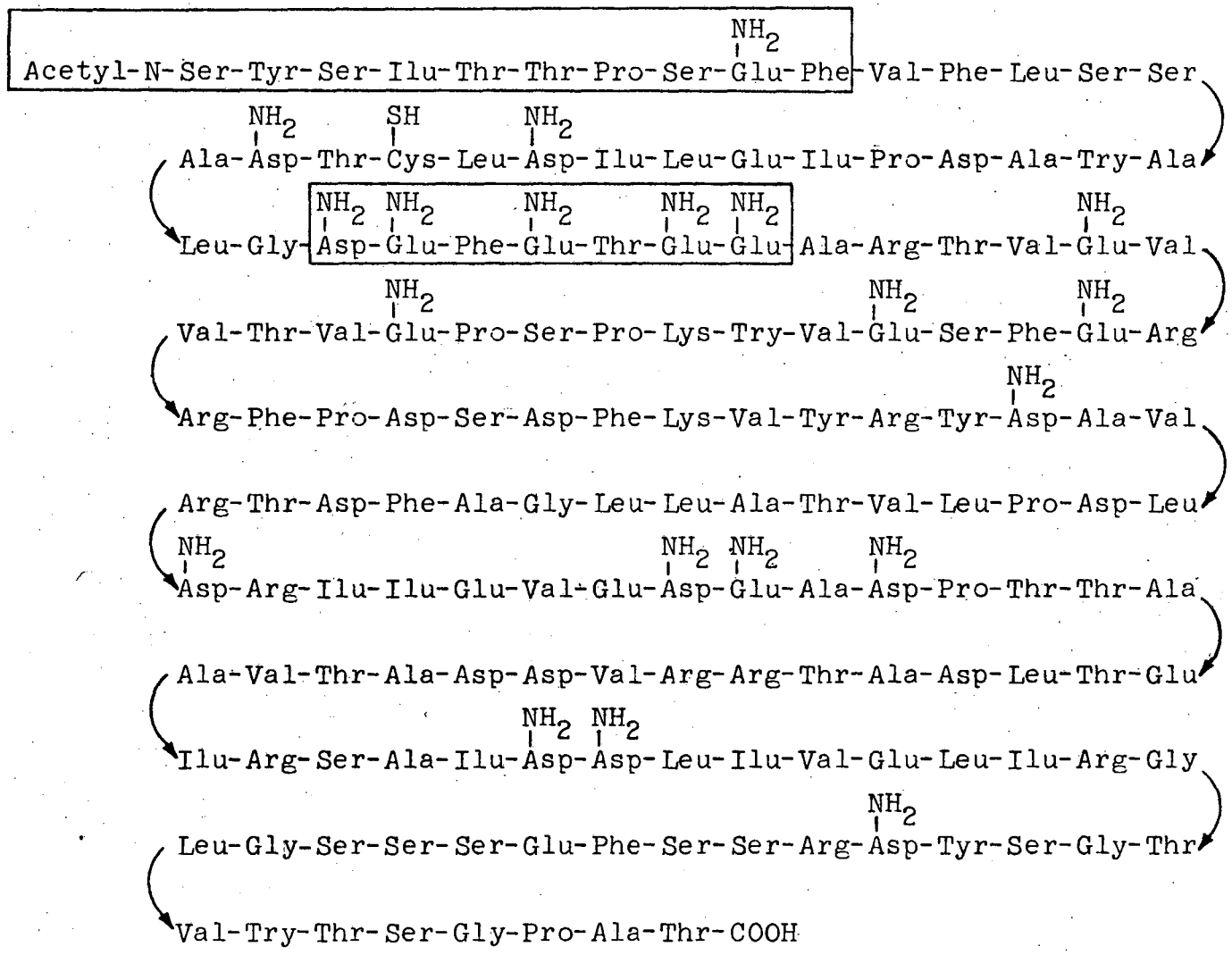
omega-mer, and since the chainlength of the omega-mer is 70 nucleotides, the guanine at the 3' end of the omega-mer is probably as little as 10 to 30 nucleotides removed from the end of the viral RNA. If the omega-mer is within 10 to 20 nucleotides of the end of the virus it should be quite feasible to determine the sequence of the intervening nucleotides. This could be done by partial hydrolysis of the partially stripped virus with T-1 ribonuclease. If the amount of stripping were kept small, one would have relatively few fragments to isolate. The result of these considerations is that the sequence of the omega-mer is of much more interest than would be the case if it were located in the interior of the RNA strand. The fortuitous position of the omega-mer makes it reasonable to suppose that one can determine the sequence of the first one hundred nucleotides of TMV-RNA using existing techniques.

In addition to locating the psi-mers and omega-mer on the TMV-RNA strand, the stripping experiments of Mandeles also support the thesis that these oligomers are unique in the viral RNA. Although the results of the pancreatic ribonuclease fingerprints are very strong evidence for the homogeneity of the three oligomer preparations, the information obtainable from a fingerprinting experiment does not provide a conclusive test of homogeneity. Consider the case of the trimers UpUpGp and CpCpGp. An equimolar

mixture of these trimers would appear homogeneous in a fingerprinting experiment performed with pancreatic ribonuclease. The fingerprint obtained would be identical with that of either of the single trimers UpCpGp or CpUpGp. However, the probability of the occurrence of such a relationship between the sequences of the two 26-mers or two 70-mers is extremely low. Another objection fo such a coincidence is that it would require the yield of oligomer obtained from the TMV-RNA to be only 35% - 40% which is unreasonably low. The stripping experiments provide a third objection. They indicate that if one were, for example, to assume the existence of two fragments in the omega-mer preparation, not only would they have to have compatible fingerprints, but they would also have to occupy nearly adjacent positions in the RNA chain. The existence of these constraints practically insures the homogeneity of the three oligomer preparations.

2. The Genetic Function of the Psi-mers and Omega-mer

Although the sequences of the psi-mers and omega-mer have not yet been obtained, one can draw some conclusions about the genetic function of the regions of the TMV-RNA which comprise these oligomers. The amino acid sequence of the coat protein of TMV-RNA is shown in Fig. 27. Due to the degeneracy of the genetic code, it is impossible to use this amino acid sequence to construct a unique gene



The boxes indicate long stretches of amino acids whose code words do not contain G.

Figure 27. The Amino Acid Sequence of Common TMV Coat Protein.[61,62]

for the coat protein. However, one can look for sequences of amino acids in the coat protein whose code words are not required to contain guanine. Long stretches of these amino acids may correspond to long oligomers in a T-1 ribonuclease digest of TMV-RNA.

There is no sequence in the TMV coat protein long enough to account for an oligomer the size of the omega-mer and there is only one such sequence which is long enough to correspond to the psi-mers. If one constructs the various models, allowed by the genetic code, of the RNA corresponding to this amino acid sequence, one sees that it is impossible to obtain a T-1 oligomer of 26 nucleotides. The closest nucleotide sequences which one can construct have chainlengths of 24, 27, 28, and 31. Furthermore, the pancreatic fingerprints of these sequences are quite different from those observed for the psi-mers. Thus, one concludes that none of the three T-1 oligomers studied in this work can be part of the gene which codes for the coat protein of TMV. This conclusion is in agreement with the very recently published findings of Kado and Knight.[63] These authors have determined that the gene for the coat protein is located in the first half of the viral RNA beginning with the 5' end of the molecule. As mentioned in Chapter VI.1, the psi-mers and omega-mer are all located in the opposite half of the RNA strand. It is unfortunate that none of the other proteins produced by the virus have even been identified.

The particularly interesting properties of the omega-mer are its unusual length and the rather severe limitations on the amino acids for which it can code. If one takes a random sequence of 6500 nucleotides, having an appropriate composition,[64] as a model for TMV-RNA, one can approximate the probability of finding a T-1 oligomer of n or more nucleotides in the following manner. If one is to obtain a fragment of chainlength greater than or equal to n from a T-1 ribonuclease hydrolysate, there must exist a sequence of at least n-1 of the nucleosides A, U, and C succeeded by G. The probability of choosing such a sequence of nucleosides from an infinite reservoir is $w(1 - w)^{n-1}$, where w is the fraction of G in the reservoir. In a molecule of N, greater than n, nucleosides there are N - n positions at which the terminal G may occur. The probability that such a sequence will appear nowhere in the chain of N nucleosides is

$$\left[1 - w(1 - w)^{n-1}\right]^{N-n}$$

Therefore, the probability that a least one such sequence will occur is given by the expression

$$1 - \left[1 - w(1 - w)^{n-1}\right]^{N-n}$$

For TMV-RNA, with w = 0.243, one computes the probability of finding one or more T-1 oligomers of chainlength greater than or equal to 70 nucleotides to be on the order of ten to the minus five.

One can look upon the course of evolution as being the result of two opposing processes. The process of mutation tends to produce random fluctuations in the sequence of an RNA while the process of selection tends to freeze those fluctuations which enhance an organism's ability to survive. The tendency of the selection process to prevent the propagation of changes in the RNA sequence depends upon the function of the RNA. Thus, an examination of the amino acid sequences of cytochromes c from various sources[65] shows that, while the sequences of certain regions of the enzymes may vary considerably, the region which binds the heme group shows very little variation in amino acid sequence. In general, one expects regions such as the active site of an enzyme to undergo less variation during the process of evolution than the enzyme as a whole. The existence of an RNA sequence of extremely low probability, as calculated on the basis of a random model, can be taken as an indication that the region of RNA involved serves a function vital to the survival of the organism. Since the omega-mer has such a low probability of occurrence, one wonders if it might not be responsible either for some control mechanism or for the synthesis of the active site of an enzyme.

Proceeding on the assumption that the omega-mer is part of one of the TMV genes, one sees that the pancreatic fingerprint places certain restrictions on the amino acids which can be encoded in this sequence. The following four

observations are significant:

1) There is a single terminal Gp.
2) A block of Ap's longer than two occurs only twice.
3) A Cp preceded by a pyrimidine occurs only once.
4) Fifty percent of the nucleotide composition is accounted for by Ap.

The first observation eliminates from consideration the eight amino acids whose code words must contain a Gp. The remaining amino acids and their non-G-containing code words are shown in Table VII. The second statement requires lysine to appear no more than twice. The appearance of lysine depends upon the phase in which the two blocks of Ap's are read. Since there are three possible phases, the probability of lysine appearing twice is 1/9. The probability of only one of the blocks being read in the correct phase is 4/9; and there is a probability of 4/9 that lysine will not appear at all. By similar reasoning, one can show from statement three that there are two chances in three that neither serine nor proline will occur. One can obtain, at most, one serine and no proline or one proline and no serine. At this point, there are nine remaining amino acids which have not been either eliminated or restricted to limited occurrence. However, due to the high adenine content of the omega-mer, at least half of the 23 amino acids being coded must come from the group: methionine, threonine, glutamine, and asparagine, which contain two Ap's in their code words.

TABLE VII

AMINO ACID CODE WORDS THAT DO NOT CONTAIN G

Code Words Containing the Sequences UC, CC, or AAA			
AAA	lysine	UCA, UCU, UCC	serine
AUC	isoleucine		
ACC	threonine		
UUC	phenylalanine	CCA, CCU, CCC	proline
CUC	leucine		
Code Words Containing Two A's			
AAU, AAC	asparagine	AUA	methionine
UAA	stop	ACA	threonine
CAA	glutamine		
Remaining Code Words			
UUU	phenylalanine	ACU	threonine
AUU	isoleucine	UAC, UAU	tyrosine
CUA, CUU, UUA	leucine	CAC, CAU	histidine

Furthermore, if the Up's occur in blocks such that they code for the maximum of four phenylalanines, the above group of four amino acids will comprise about 16 of the remaining 19 possibilities.

One expects that enzymes which have similar or identical specificities will have similar amino acid sequences in various species.[65,66,67] Thus, one might be able to determine what type of enzyme is made at the 3' end of the viral RNA by looking for a class of enzymes which contain a region rich in the four amino acids listed above. Unfortunately, the extensive listing of amino acid sequences needed for such an approach is not available.

The interesting feature of the psi-mers is the similarity of their fingerprints. It is currently believed that, as complex organisms evolve from simple organisms, a process of gene doubling must occur.[68] As the evolutionary process proceeds the pair of genes thus produced will gradually become dissimilar through a process of random mutation. However, if mutations of a certain region are usually lethal, a certain amount of homology might persist between the gene pair. The probability, on the basis of a random model, of finding two distinct regions of TMV-RNA which are identical to 26 nucleotides is so small (on the order of 10^{-15}) that one would consider this situation to be a negation of the random hypothesis. The obvious alternative hypothesis would be that one is observing a situation which is the result of gene doubling.

One should keep in mind the fact that, no matter how low the probability, one can not ignore the random hypothesis with complete certainty.

The two psi-mers have identical chainlengths, and, with the exception of three base changes, their fingerprints match perfectly.

C, C, AAAC, C, C, AC, AC, U, U, AU, AU, AAU, AAAU, G psi-1

U, U, U, AAC, C, C, AC, AC, U, U, AU, AU, AAU, AAAU, G psi-2

Two of the base changes are C to U conversions which is a well-known type of point mutation. The third change would require converting A to U. One would like to know whether the similarity of the above fingerprints represents an event of low enough probability to make its occurrence on a random basis doubtful.

The probability of the observed similarity in the psi-mer fingerprints is the product of the following three probability calculations:

1) The probability of obtaining a T-1 ribonuclease fragment of chainlength n when w is the fraction of G in an RNA of chainlength N:

$$P(n) = 1 - \left[1 - w^2(1 - w)^{n-1}\right]^{N-n}$$

2) The probability of obtaining a given base composition, $A_rU_sC_t$, correct to within m bases when the reservoir has the composition $A_xU_yC_z$, where r, s, and t are integers such that $r + s + t = n - 1$; and $x + y + z = 1$:

$$P(m,r,s,t) = \sum_{i,j} \frac{(r+s+t)!\ x^{r+i}\ y^{s+j}\ z^{t-i-j}}{(r+i)!\ (s+j)!\ (t-i-j)!}$$

The summation ranges over values of i and j between -m and +m such that i + j is always between -m and +m (see Table VIII).

3) The probability of obtaining a given pancreatic fingerprint from an oligomer of composition, $A_rU_sC_tG$:

$$P(fp) = \left[\sum_{h=0}^{r} \sum_{k=0}^{r-h} p_{1,h}\ p_{s,k}\ p_{t,r-h-k} \right]^{-1}$$

The $p_{i,j}$ are obtained from the array in Table IX. This last problem is equivalent to asking for the number of ways of arranging r indistinguishable objects (the A's) among three classes of containers (s U's, t C's and the terminal G), such that the classes are distinguishable, but the members of each class are not.

For the case in point, the three probabilities P(n), P(m,r,s,t), and P(fp) were computed to be 0.26, 0.47, and 0.00034 respectively. Thus the odds of obtaining the observed similarities for the psi-mers are about one in twenty thousand. These odds are unfavorable enough to give credence to the possibility that one is observing a case of gene doubling. This hypothesis could be either greatly strengthened or disproved by determing the complete sequences of both psi-mers. For this reason, one is

TABLE VIII

Values of $\frac{r!\ s!\ t!}{(r+i)!\ (s+j)!\ (t-i-j)!}\ x^i\ y^j\ z^{-i-j}$

with r = 12, s = 6, t = 7, x = 0.381, y = 0.365, z = 0.254

for i = -3····+3, j = -3····+3, such that (i+j) = -3·····+3

	i = -3	-2	-1	0	+1	+2	+3
j = +3	2.302	4.832	2.143	1.237			
+2	1.800	2.164	2.066	1.550	0.892		
+1	1.113	1.504	1.642	1.437	0.996	0.533	
0	0.543	0.814	1.000	1.000	0.807	0.520	0.260
-1		0.340	0.464	0.521	0.482	0.362	0.217
-2			0.162	0.202	0.209	0.180	0.126
-3				0.056	0.065	0.620	0.050

The sum of the values in the table, when multiplied by $\frac{(r+s+t)!\ x^r\ y^s\ z^t}{r!\ s!\ t!}$ equal to 0.01346, gives P(m,r,s,t).

TABLE IX

THE NUMBER OF WAYS OF ARRANGING N INDISTINGUISHABLE OBJECTS IN M INDISTINGUISHABLE BOXES

M \ N	0	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	2	2	3	3	4	4	5	5	6	6	7
3	1	1	2	3	4	5	7	8	10	12	14	16	19
4	1	1	2	3	5	6	9	11	15	18	23	27	34
5	1	1	2	3	5	7	10	13	18	23	30	37	47
6	1	1	2	3	5	7	11	14	20	26	35	44	58
7	1	1	2	3	5	7	11	15	21	28	38	49	65
8	1	1	2	3	5	7	11	15	22	29	40	52	70
9	1	1	2	3	5	7	11	15	22	30	41	54	73
10	1	1	2	3	5	7	11	15	22	30	42	55	75
11	1	1	2	3	5	7	11	15	22	30	42	56	76
12	1	1	2	3	5	7	11	15	22	30	42	56	77

particularly interested in sequencing schemes which might be feasible for the psi-mers. Two such schemes are discussed in the following section.

3. Possible Methods for Obtaining Complete Sequences for the Psi-mers and Omega-mer

The main difficulty in the sequencing of the psi-mers and omega-mer is that the overlap technique, which has been a great aid in the sequencing of the transfer-RNAs, cannot be applied. A possible technique which one might use in place of the overlap method is a combination of two sequencing schemes which have been suggested by other authors.

Mandeles and Tinoco[69] suggested the use of a label at one or both ends of the RNA in conjunction with partial hydrolysis by T-1 or pancreatic ribonuclease. One can use the label to pick out a hierarchy of fragments in which the terminal label is associated with fragments of increasing chainlength. Be determing the T-1 or pancreatic ribonuclease fingerprints of each member of the hierarchy, one can determine the complete sequence of the original RNA. The applicability of this technique is limited by one's ability to separate the fragments resulting from the partial digestion. The difficulty of this separation problem is increased by the presence of a great many fragments from the middle of the original molecule, which do not contain the label, but must be separated from those

fragments which are labeled.

The method which Holley[70] has used for oligomers with chainlengths on the order of ten avoids the above difficulty by producing the desired hierarchy by partial digestion with an exonuclease. For example, partial digestion with spleen phosphodiesterase, which attacks the terminus with a free 5' hydroxyl group, would produce a hierarchy of fragments beginning with the 3' end. The rest of the oligomer is reduced to mononucleotides. However, the fragments in the hierarchy now increase in chainlength by only one nucleotide at a time; and, therefore, the hierarchy contains a larger number of fragments than would be the case if a base-specific endonuclease had been used. Thus, there is an increase in the number of fragments one must examine; and a corresponding decrease in the yield of each fragment.

One can combine the two techniques described above in the following manner. The recently discovered 5' phosphokinase[71] allows one to add a phosphate group to the 5' end of the oligomer to be sequenced. A partial hydrolysis of this oligomer with a base-specific endonuclease, such as pancreatic ribonuclease, will now result in fragments containing a free 5' hydroxyl group except for fragments coming from the 5' end of the oligomer, which will contain a terminal 5' phosphate. The partial hydrolysis can be terminated by removing the nuclease with bentonite. Further treatment of the fragments with spleen

phosphodiesterase will cause those pieces containing the free 5' hydroxyl group to be degraded to mononucleotides. However, the presence of a terminal phosphate has been shown to seriously inhibit the action of the diesterase.[72] Thus, the final digestion mixture will contain only mononucleotides and a hierarchy of fragments containing the 5' terminal end of the original oligomer. Furthermore, the hierarchy will contain fewer fragments than would have been produced by the action of the diesterase alone.

The method outlined above will be applicable for chainlengths up to the limit of the resolving power of the separation technique used. For long chainlengths it would be necessary to combine two or three partial digests, performed under different conditions, in order to obtain an even distribution of chainlengths in the hierarchy. The chainlength separation chromatography described in Chapter II.3 can be greatly improved by increasing the column height and decreasing the sample load. It should be feasible to separate fragments according to chainlength up to about 30 nucleotides, especially if most of the fragments differ in chainlength by two or more nucleotides. Thus the scheme described above should be a practical way of determining the sequence of the psi-mers. If the acid soluble nuclease discussed in Chapter V.2 were used to produce the hierarchy, it would greatly simplify the chainlength separation by increasing the average difference in chainlength between neighboring members of

the heirarchy. One needs a minimum of 0.02 μM of each fragment in the hierarchy in order to obtain its fingerprint by the method used in this work. Since each of the psi-mers yields roughly 15 fragments after digestion with pancreatic ribonuclease, approximately 0.3 μM or 3 mg, of each psi-mer would be required to perform a single sequence determination. This estimate is an absolute minimum. One could work more comfortably with amounts of material two to four times larger. The omega-mer, unfortunately, lies beyond the scope of this method with the currently available separation techniques.

As mentioned in the last section of the preceding chapter, the problem of sequencing a long oligomer is greatly simplified if a method can be found of selectively breaking the oligomer into two or three large pieces. The approach used in Chapter V was not successful; but Sanger[4] has reported that, at high concentrations, T-1 ribonuclease can be made to cleave on the 5' linked side of adenosine. This observation might offer a way of obtaining a small number of specific breaks in the omega-mer, particularly if T-1 ribonuclease behaves in a manner similar to that shown by pancreatic ribonuclease. In Chapter IV it was mentioned that the secondary specificity of pancreatic ribonuclease is such that it tends to cleave between two adjacent adenosine residues. If T-1 ribonuclease exhibits a similar sequence preference in its secondary activity one would be able to obtain

specific fragments from the omega-mer in large enough yields to allow further work on the individual fragments. For example, suppose one were able to cleave the omega-mer into three fragments of chainlength 20 to 30 with only 75% efficiency. About 0.4 μM, or 8 mg, of omega-mer would be required to yield a sufficient amount of each of the fragments to allow their sequences to be determined by the method of partial digestion discussed above.

Despite the difficulties involved in sequencing T-1 ribonuclease fragments the size of the psi-mers and omega-mer, it is the opinion of this author that the problem can be solved with currently available technqiues, and that there is considerable incentive for continuing this work. The pancreatic ribonuclease fingerprints of these oligomers give only a portion of the sequence; but they provide a solid foundation for further studies in that they supply information about the number and type of fragments which one can expect to obtain from partial hydrolysates. This information should be very useful in the interpretation of further experiments.

The intriguing similarity of the psi-mer fingerprints evokes considerable interest in the sequences of both of these oligomers; and the stripping experiments of Mandeles indicate that the sequencing of the omega-mer will allow a large portion of the terminal sequence of TMV-RNA to be determined.

APPENDIX I

Ion Exchange Chromatography - Theory

The theory of ion-exchange chromatography has been treated by a number of authors.[13,14,73] None of these theories discusses the situation in which elution is accomplished with a concentration gradient of the carrier ion, a technique which is quite important in biological work. The theory to be developed in this section is primarily concerned with the treatment of gradient elution for ion-exchange chromatography. However, the initially derived expressions will also apply to systems involving both ion-exchange chromatography and gel filtration. The goal of the development will be to derive a theoretical relationship to explain the behavior of the chainlength separation technique described in Chapter II.3. The theory is also useful in that it illustrates the dependence of the resolution obtained from a chromatographic system on various experimental parameters such as the flowrate, the size of the column, and the resin capacity. The theory can be used qualitatively to judge the direction in which these parameters must be varied in order to increase resolution; or it can be used quantitatively. A trial column can be used to evaluate the undetermined constants in the theory and one can then calculate the optimum values of all experimentally controllable parameters.

The most comprehensive treatment of ion-exchange chromatography appears to be that of Vermeulen and Hiester.[14] This work will follow both the notation and form of their development fairly closely. It will be shown that with only a few modifications the basic expressions derived by these authors can be applied to the case of gradient elution. The underlying assumption of the theory published by Vermeulen and Hiester is that the chromatography is performed under trace conditions. By trace conditions one means that the concentrations of the sample species, both in the solid and liquid phases, are small compared to the concentrations of the carrier ion. This assumption allows one to neglect fluctuations in the carrier ion concentration caused by the interaction of the sample species with the resin. It also allows one to neglect any interaction between different sample species. Thus, each component can be treated independently. In practice, one finds that better resolution is obtained if the load placed on a column is kept small. The question of whether the usual sample loads, which one employs, fit the assumption of trace conditions will be considered later in order to determine whether this assumption seriously limits the theory.

The rate at which a component equilibrates between the resin and the solvent on a chromatographic column may be limited by any one of a number of mechanisms. In the majority of cases the limiting process will be the diffusion

of molecules from the mobile solvent phase into the layer of solvent immediately surrounding the resin particles, or, in the case of a porous resin, into the resin itself. The actual ion-exchange process will generally be quite rapid, as in the case with most ionic reactions. It should be pointed out that Vermeulen and Hiester[14] have shown that all of the mechanisms which one is likely to encounter lead to rate equations of the same form in the limit of trace conditions. Thus, one may focus attention on the diffusion limited case without loss of generality.

Consider the competition between a polyanion A^{-n} and a monovalent carrier ion B^{-1} for sites on an anion-exchange resin,

$$A^{-n} + n(B\cdot resin) \rightleftharpoons B^{-1} + (n-1)(B\cdot resin) + (A\cdot resin)^{-(n-i)} \rightleftharpoons \text{---} \; nB^{-1} + (A\cdot resin) \qquad (1)$$

The concentration of $(A\cdot resin)^{-(n-i)}$ in meq. per gm of dry resin will be denoted by $q_{A,i}$. Similarly, the concentration of carrier ion on the resin will be denoted by q_B. The solution concentrations of the sample and carrier ion in meq. per ml will be defined as c_A and c_B. One can now define a set of equilibrium constants.

$$K_{A,i} = \frac{q_{A,i}\, c_B^*}{c_A^*\, q_B} \quad ; \quad K_{A,i} = \frac{q_{A,i}\, c_B^*}{q_{A,i-1} q_B} \quad ; \quad i = 2, \text{---} n \,, \qquad (2)$$

where the asterisk refers to solution concentrations in

equilibrium with the resin. One wishes to obtain a relation between the total concentration of A^{-n} bound to the resin, q_A, and the solution concentration, c_A^*. Such a relation takes the following form:

$$q_A = \sum_{i=1}^{n} q_{A,i} = c_A^* \sum_{i=1}^{n} \left(\frac{q_B}{c_B^*}\right)^i \prod_{j=1}^{i} K_{A,j} \quad . \tag{3}$$

At this point one can invoke the assumption of trace conditions to simplify the above expression. Since q_A/q_B will be very small one can set q_B equal to Q, the total capacity of the resin available to the polyanion. In addition c_A^*/c_B^* will be very small so that c_B^* can be set equal to c_B, the carrier ion concentration in the bulk of the eluent. For the case of gradient elution, c_B will be some predetermined function of the eluent volume. In practice the rate of increase of the carrier ion concentration during gradient elution will be extremely slow compared to the rate at which the solvent attains equilibrium with the resin. Thus, the resin may be considered to always be in equilibrium with the carrier ion; and the approximation, c_B^* equals c_B, will remain valid. With the use of these assumptions Equation (3) can be simplified to

$$q_A = c_A^* \left(\frac{Q}{c_B}\right)^n \sum_{i=1}^{n} \left(\frac{c_B}{Q}\right)^{n-i} \prod_{j=1}^{i} K_{A,j}$$

$$= c_A^* \left(\frac{Q}{c_B}\right)^n K_A \quad . \tag{4}$$

This expression defines an effective equilibrium constant which is a function of c_B, and, thus, is indirectly a function of the elution volume. Later, a simplified model for the reaction of the polyanion with the resin will be introduced to remove the dependence of K_A upon the elution volume. However, this is not necessary for the development of the theory.

One is now in a position to write the rate equation governing the approach to equilibrium between the solvent phase and the resin. The rate at which this equilibrium is attained will be limited by the diffusion of the sample component between the moving solvent and that portion of the solvent, in and around the resin particles, which is in equilibrium with the resin. The rate of increase with time in the number of moles of a species contained on the resin and in the solvent which is in equilibrium with the resin must equal the diffusional flow of that species out of the external solvent. Assuming the concentration gradient between the external solvent and the layer of solvent adjacent to the resin to be constant, one can write the following rate equation,

$$\frac{d(\rho q_A)}{dt} + \frac{d(f_i c_A^*)}{dt} = D_A \frac{\alpha}{\delta} (c_A - c_A^*) \quad , \qquad (5)$$

where ρ is the resin density in gm of dry resin per ml of total bed volume and f_i is the fraction of the bed volume

occupied by the solvent which is in equilibrium with the resin and also accessible to the sample species. The bed volume itself, V_b, is defined to be the total volume in ml of the packed column. D_A is the diffusion coefficient of species A^{-n} with units of $cm^2/sec.$, α is an effective area in cm^2 per ml of bed volume, and δ is the effective distance in cm over which diffusion must occur.

It is desirable to rewrite Equation (5) in a different form. First, one can express the time, t, in terms of the eluent volume, V_e, and the void volume, $f_e \cdot V_b$, according to the equation,

$$V_e = Rt - f_e V_b , \tag{6}$$

where R is the flowrate of the column in ml/sec. This equation gives the volume of eluent which has passed through a given bed volume at time t, assuming that the eluent just entered the top of the column at t=0. One would also like to express Equation (5) in terms of dimensionless parameters. This can be partially accomplished by dividing both sides of the equation by some arbitrary standard concentration c_A^o. The expression

$$q_A^o = c_A^o K_A^o \left(\frac{Q}{c_B^o}\right)^n \tag{7}$$

relates c_A^o to a standard resin concentration q_A^o, a standard carrier ion concentration c_B^o, and the associated equilibrium

constant K_A^o. The relations (4), (6), and (7) allow one to put Equation (5) into the form

$$\left(\frac{\partial}{\partial V_e}\left[f_i\frac{c_A^*}{c_A^o}+\rho K_A^o\left(\frac{Q}{c_B^o}\right)^n\frac{q_A}{q_A^o}\right]\right)_{V_b}= \frac{D_A}{R}\frac{\alpha}{\delta}\left[\frac{c_A}{c_A^o}-\frac{\left(f_i\frac{c_A^*}{c_A^o}+\rho K_A^o\left(\frac{Q}{c_B^o}\right)^n\frac{q_A}{q_A^o}\right)}{f_i+\rho K_A\left(\frac{Q}{c_B}\right)^n}\right]. \tag{8}$$

The terms involving f_i in Equation (8) will give rise to a gel-filtration effect. By setting K_A equal to zero one can treat the process of gel-filtration alone, or if one includes a distribution coefficient in f_i, one can treat the process of partition chromatography. For nonporous resins, such as polystyrene derivatives, f_i will always be negligible and no gel-filtration will occur. In the case of porous resins, such as crosslinked dextran derivatives, f_i will be much larger; and gel-filtration will be possible when $\rho K_A Q^n/c_B{}^n$ is less than one. However, in most instances $\rho K_A Q^n/c_B{}^n$ will be much larger than one, and ion-exchange will predominate. Although gel-filtration will not be discussed in detail in this work, the terms involving f_i will be retained in order to illustrate how the process of gel-filtration can be incorporated into the theory.

One ultimately wants to solve Equation (8) for c_A/c_A^o in terms of the eluent volume and the bed volume. However, there is a constraint on the system which must be considered. This constraint takes the form of a conservation equation which states that the number of equivalents of a species which are lost from a volume of eluent must be gained by the resin through which it has passed,

$$-\left(\frac{\partial\, c_A/c_A^o}{\partial\, V_b}\right)_{V_e} = \left(\frac{\partial}{\partial V_e}\left[f_i\,\frac{c_A^*}{c_A^o} + \rho K_A^o\left(\frac{Q}{c_B^o}\right)^n \frac{q_A}{q_A^o}\right]\right)_{V_b} . \qquad (9)$$

Equation (9) requires that any solution to the rate equation must be a function $F(V_b,V_e)$ such that

$$\left(\frac{\partial F}{\partial V_e}\right)_{V_b} = -\,\frac{c_A}{c_A^o} \quad ; \quad \left(\frac{\partial F}{\partial V_b}\right)_{V_e} = f_i\,\frac{c_A^*}{c_A^o} + \rho K_A^o\left(\frac{Q}{c_B^o}\right)^n \frac{q_A}{q_A^o} . \qquad (10)$$

By substituting (10) into (9) and making the transformations

$$x_A = \frac{D_A}{R} \frac{\alpha}{\delta} \int_0^{V_e} \frac{dz}{f_i + \rho K_A(z) \left(\frac{Q}{c_B(z)}\right)^n} \quad ; \tag{11}$$

$$y_A = \frac{D_A}{R} \frac{\alpha}{\delta} V_b \; ; \; F(x_A, y_A) = e^{-(x_A + y_A)} \phi(x_A, y_A)$$

one obtains the differential equation

$$\frac{\partial^2 \phi}{\partial x_A \partial y_A} - \phi = 0 \tag{12}$$

This equation has the same form as that obtained by Vermeulen and Hiester[14] and has been solved for the appropriate boundary conditions by Thomas.[73] One first solves the equation for the boundary conditions applicable for the loading of the column,

$$\left(\frac{\partial F}{\partial y_A}\right)_{x_A=0} = 0 \; ; \left(\frac{\partial F}{\partial x_A}\right)_{y_A=0} = 1 \tag{13}$$

where the standard concentrations c_A^o and c_B^o have been chosen to be the concentrations utilized during the loading of

the column. The boundary conditions in (13) allow one to determine the distribution of sample on the resin at the end of the loading operation,

$$\left(\frac{\partial F}{\partial y_A}\right)_{x_A=x_A'} = \int_0^{x_A'} e^{-(z + y_A)} I_o \left(2\sqrt{zy_A}\right) dz \tag{14}$$

$$x_A' = \frac{D_A}{R} \frac{\alpha}{\delta} \frac{V_e'}{f_i + \rho K_A^o \left(\frac{Q}{c_B^o}\right)^n} = \frac{D_A}{R} \frac{\alpha}{\delta} \frac{N_A}{f_i c_A^o + \rho q_A^o}$$

where I_o is the modified Bessel function of zeroth order. Primed quantities, such as x_A', refer to values at the end of the loading operation; and N_A is the number of equivalents of A^{-n} placed on the column.

The solution for the loading operation can now be used as a boundary condition for the elution process. The appropriate boundary conditions for elution are

$$\left(\frac{\partial F}{\partial y_A}\right)_{x_A=0} = \int_0^{x_A'} e^{-(z + y_A)} I_o \left(2\sqrt{zy_A}\right) dz \quad ;$$

$$\left(\frac{\partial F}{\partial x_A}\right)_{y_A=0} = 0 \quad . \tag{15}$$

For these boundary conditions one obtains -

$$-\left(\frac{\partial F}{\partial x_A}\right)_{y_A} = \frac{c_A}{c_A^o} = \int_o^{y_A} e^{-(z + x_A)} I_o\left(2\sqrt{zx_A}\right)dz$$

$$-\int_o^{y_A} e^{-(z + x_A + x_A')} I_o\left(2\sqrt{z(x_A + x_A')}\right)dz \quad . \tag{16}$$

This expression is not particularly suitable for either physical interpretation or simple numerical calculations. Fortunately, the above expression can be closely approximated by a simple error function.[74] Vermeulen and Hiester[14] have shown that, if one assumes y_A to be large compared to one, and x_A' to be much smaller than y_A, the following approximation can be made in the vicinity of $x_A = y_A$:

$$\frac{c_A}{c_A^o} \simeq \frac{x_A'}{2\sqrt{\pi y_A}} e^{-[x_A - (y_A - \frac{1}{2} x_A')]^2/4y_A} \quad . \tag{17}$$

The gaussian form shown above will be a valid approximation of the band shape when x_A is within $y_A^{\frac{1}{2}}$ of the maximum, but will tend to over-estimate the leading edge of the band and under-estimate the trailing edge of

the band. A detailed analysis of the error involved in this approximation is given by the above authors. The assumption of small x'_A is equivalent to the assumption that the number of equivalents of sample loaded on the column is small compared to the maximum load for the given conditions, c^o_A and c^o_B. The assumption of large y_A requires that the rate at which the solvent equilibrates with resin be large compared to the time required for the passage of solvent through the column. Both of these assumptions correspond to empirically derived conditions required for good resolution.

One is now in a position to set up some sort of experimental criteria for deciding whether a particular chromatographic system fits the assumption of trace conditions. Equation (17) predicts that, if all other parameters are constant, the concentration at the band maximum should be directly proportional to the sample load placed on the column. One expects that as the trace condition assumption breaks down that a plot of concentration at the band maximum versus sample load, for a given species, should begin to level off. The reason for expecting this behavior is that, as one moves out of the region of trace conditions, one will begin to saturate the resin. This, in turn, will cause the band to be broader than the theory would predict. Thus, a linear plot of the type described above would serve as an experimental verification of the

assumption of trace conditions. Equation (17) also indicates that, even if the column is not loaded under trace conditions, one can remove the effect of the non-trace behavior. This can be done by making the bed volume of the column large enough so that x_A' is completely negligible compared to y_A. When x_A' is sufficiently small (less than one-hundredth the value of $y_A^{\frac{1}{2}}$) one can make a further approximation by ignoring x_A' in the exponent. This approximation will be assumed in the discussion that follows.

Having obtained an expression for the behavior of an ion-exchange column, one is in a position to investigate the effect of various parameters on the resolution. For the purposes of this discussion it is desirable to eliminate some of the generality which has been retained up to this point. The effect of gel-filtration will be ignored as well as the dependence of K_A upon the carrier ion concentration. It will be shown later that in most circumstances K_A may be considered to be constant. The carrier ion concentration will be considered to be a linear function of V_e having the form,

$$c_B(V_e) = gV_e \quad . \qquad (18)$$

The variable x_A will now have the form

$$x_A = \frac{D_A}{R} \frac{\alpha}{\delta} \frac{g^n}{n+1} \frac{V_e^{n+1}}{\rho K_A Q^n} \quad . \tag{19}$$

In the vicinity of the maximum one can approximate x_A with an expression which is linear in V_e,

$$x_A = \frac{D_A}{R} \frac{\alpha}{\delta} \left(\frac{g^n \bar{V}_e^n}{\rho K_A Q^n} V_e - n V_b \right) \tag{20}$$

$$\bar{V}_e = \left[\frac{(n+1)\rho K_A Q^n}{g^n} V_b \right]^{\frac{1}{n+1}} ,$$

where the bar will be used to denote values at the maximum.

The amount of resolution obtained for two species which elute in gaussian shaped bands of the general form,

$$\frac{c}{2\sqrt{\pi b}} e^{-(x-a)^2/4b} \tag{21}$$

will be defined to be

$$\Delta = \frac{1}{2} \frac{a_2 - a_1}{\sqrt{b_2} + \sqrt{b_1}} \quad . \tag{22}$$

Thus when the resolution is less than one, a single band will be visible; and as the resolution becomes larger than one, two peaks will appear. By substituting values for y_A and x_A from Equations (13) and (22) respectively into Equation (19) one obtains $\Delta =$

$$\frac{1}{2}\left(\frac{V_b}{R}\,\frac{\alpha}{\delta}\right)^{\frac{1}{2}} \frac{\left[\frac{(n+1)Q^nK_2}{g^n}\right]^{\frac{1}{n+1}} - \left[\frac{(m+1)Q^mK_1}{g^m}\right]^{\frac{1}{m+1}}}{\frac{1}{(n+1)\sqrt{D_2}}\left[\frac{(n+1)Q^nK_2}{g^n}\right]^{\frac{1}{n+1}} + \frac{1}{(m+1)\sqrt{D_1}}\left[\frac{(m+1)Q^mK_1}{g^m}\right]^{\frac{1}{m+1}}} \qquad (23)$$

where m and n are the charges possessed by species one and two respectively. For species having a single charge, the above expression reduces to

$$\Delta = \left(\frac{V_b}{R}\,\frac{\alpha}{\delta}\right)^{\frac{1}{2}} \frac{\sqrt{K_2} - \sqrt{K_1}}{\sqrt{\frac{K_2}{D_2}} + \sqrt{\frac{K_1}{D_1}}} \quad . \qquad (24)$$

Thus, the resolution should improve without limit as the bed volume is increased. The optimum bed volume will only be determined by the largest volume with which one can conveniently work. One should note that the height

to diameter ratio of the column does not enter into the theory. This result is contingent upon the implicit assumption that the exact nature of the flow through the column can be ignored. This only only be true if one avoids the extremes of very narrow and very short columns. The size of the resin particles will enter the expression through α and δ. One would expect α/δ to be roughly proportional to the inverse square of the particle diameter. Thus, decreasing the particle size should improve resolution. One cannot predict the effect of temperature unless one knows the temperature dependence of K and D. For identically charged species the carrier ion gradient and the resin capacity should not effect the resolution. For species having different charges, the resolution should increase with increasing resin capacity and decreasing carrier ion gradient. Equations (23) and (24) predict that the resolution will increase indefinitely with decreasing flowrate. This prediction ignores the tendency of a band to broaden by a process of lateral diffusion.

The exact treatment of the lateral diffusion effect leads to a higher order equation whose solution is not apparent. However, one may treat this effect in an approximate fashion by allowing the lateral diffusion to occur independently of the chromatography and by considering the gaussian shaped band to be fully formed at the beginning of the elution process. It is difficult to assess the

accuracy of these approximations, but they should be satisfactory for obtaining a rough estimate of the region in which lateral diffusion becomes a serious problem. One takes advantage of the fact that a gaussian shaped band will retain its general form during the process of diffusion. Thus, the differential equation and associated boundary condition,

$$\left(\frac{\partial f(x,\tau)}{\partial \tau}\right)_x = \left(\frac{\partial^2 f(x,\tau)}{\partial x^2}\right)_\tau ; \quad f(x,o) = \frac{c}{2\sqrt{\pi b}} e^{-(x-a)^2/4b} \qquad (25)$$

will have the solution

$$f(x,\tau) = \frac{c}{2\sqrt{\pi(b+\tau)}} e^{-(x-a)^2/4(b+\tau)} . \qquad (26)$$

The equation governing lateral diffusion will be

$$\left(\frac{\partial (c_A/c_A^o)}{\partial t}\right)_{V_e} = D_A (Sf_e)^2 \left(\frac{\partial^2 (c_A/c_A^o)}{\partial V_e^2}\right)_t \qquad (27)$$

where S is the cross sectional area of the column. In physical terms one wishes to first perform the chromatography and then allow diffusion to occur for a period equal to the

sample retention time. Thus Equation (6) which couples the variables V_e and t will no longer apply. V_e will now only be a measure of the elution volume in the vicinity of the band maximum. One can apply the following transformations to bring Equation (27) into the form shown in (25):

$$x_A = \frac{D_A}{R} \frac{\alpha}{\delta} \left(\frac{g^n \bar{V}_e^{\,n}}{\rho K_A Q^n} V_e - nV_b \right)$$

$$\tau_A = D_A = \frac{D_A}{R} \frac{\alpha}{\delta} \left(\frac{Sf_e g^n \bar{V}_e^{\,n}}{\rho K_A Q^n} \right)^2 \bar{t} \simeq \frac{D_A}{R\bar{V}_e} \left(\frac{D_A}{R} \frac{\alpha}{\delta} Sf_e V_b \right)^2 . \quad (28)$$

Thus, one obtains

$$\frac{c_A}{c_A^o} = \frac{x_A'}{2\sqrt{\pi(y_A + \tau_A)}} e^{-(x_A - y_A)^2/4(y_A+\tau_A)} . \quad (29)$$

The exponent in Equation (29) can be expressed in the following form by substituting values for x_A and τ_A from Equation (28):

$$\frac{(x_A-y_A)^2}{4(y_A+\tau_A)} = \frac{\left(\frac{n+1}{2}\right)^2 \frac{D_A}{R} \frac{\alpha}{\delta} \frac{V_b}{\bar{V}_e}}{\bar{V}_e + \frac{\alpha}{\delta} V_b \left(Sf_e \frac{D_A}{R}\right)^2} (V_e - \bar{V}_e)^2 \quad . \qquad (30)$$

Although the resolution will initially vary inversely with the square root of R, in the region in which lateral diffusion predominates any further reduction in flowrate will cause a proportional decrease in resolution. With the inclusion of the lateral diffusion effect, the height to diameter ratio of the column becomes important and the resin capacity and carrier ion gradient affect the resolution even in the absence of a charge difference between two species.

When lateral diffusion is negligible, one can calculate K_A from the position of the band maximum. The value of the half width will allow one to calculate $D_A\ \alpha/\delta$. The values of various parameters applicable for a chainlength separation similar to that shown in Figure 1 of Chapter II.3 are given in Table X. These values are intended only as rough estimates to give one a feeling for the magnitudes of various parameters. The values of $D_A\ \alpha/\delta$ were calculated by ignoring lateral diffusion. One should note that some of the bands contain more than one species. While this should not greatly affect the position of the maximum, it

TABLE X

CHROMATOGRAPHIC PARAMETERS FOR CHAINLENGTH SEPARATION ON DEAE-SEPHADEX, A-25

Units / Sample	Negative Electronic Charges	ml.	ml./sec	$\frac{meq}{ml^2}$	gm/ml.	meq./gm
	N	V_b [a]	R [a]	g [a]	ρ [b]	Q [b]
Gp	2	100	0.011	10^{-4}	0.1	3.5
ApGp UpGp CpGp	3	"	"	"	"	"
psi-1-mer psi-2-mer	27	"	"	"	"	? [c]
omega-mer	71	"	"	"	"	? [c]

ml.	gm/ml.	gm/ml.	$\left(\frac{gm}{ml}\right)^n$	sec^{-1}	unitless	ml/sec.
$\bar{V}_e$ [c]	ρK_1 [c]	k [c]	ρK_A [c]	$DA\frac{\alpha}{\delta}$ [c]	y_A [c]	R_{opt}
1240	0.04	0.13	5.2×10^{-3}	330×10^{-4}	300	4×10^{-5}
1840	"	"	6.6×10^{-4}	140×10^{-4}	130	1×10^{-5}
4360	"	"	?	41×10^{-4}	37	3×10^{-6}
4760	"	"	?	6.9×10^{-4}	6.2	4×10^{-7}

[a] Independently evaluated parameters.

[b] Resin parameters obtained from Pharmacia literature.

[c] Evaluated from the elution profile.

may drastically alter the bandwidth.

The effect of lateral diffusion is primarily of interest because it allows one to determine the optimum flowrate for a chromatographic system. From Equation (30) one can determine the optimum flowrate for fixed values of the other parameters,

$$R_{opt} = D_A S f_e \left(\frac{\alpha V_b}{\delta \bar{V}_e} \right)^{1/2} . \qquad (31)$$

The data obtainable from a single chromatography experiment is not sufficient to determine the optimum flowrate unless one has an independent method of determining either D_A or α/δ. In order to determine R_{opt} from chromatographic data it is necessary to run at least two columns at different flowrates.

Let us now investigate the possibility of improving the resolution obtained in the chainlength separation described in Table X . One can always increase the bed volume, but this is not a particularly interesting alternative since this parameter has no optimum value. Thus, the bed volume will arbitrarily be fixed at 100 ml although it is not too inconvenient to work with bed volumes 2 to 5 times larger. The other parameter which one can vary is the flowrate. In this case one wishes to make

the lateral diffusion term as small as possible. This can be done both by decreasing the cross-sectional area, and by decreasing the carrier ion gradient in order to increase $\bar{V}_e$. Since it is inconvenient to work with a column more than two meters long, 0.5 cm will be a practical lower limit for S. As one decreases the carrier ion gradient, the bandwidth becomes larger at about the same rate as the peak to peak separation increases. Thus, the resolution will not be affected, but the volume of sample collected will become inconveniently large. Therefore, the value of g will be left unchanged although one could probably decrease it by a factor of 2 or 3 without making the dilution problem too serious. One should note that reducing S and g does not affect the resolution directly, but, rather allows one to lower the optimum flowrate as much as possible.

One is now in a position to calculate the optimum flowrate. To do this it is necessary to have at least an order of magnitude estimate of α/δ. Assuming the resin particles to be spheres, α can be taken as the sum of the areas of the spheres per unit bed volume,

$$\alpha = N\,4\pi r^2 = \frac{3(1-f_e)}{r} , \qquad (32)$$

where r is the average radii of the resin particles and N

is the number of particles per unit bed volume. The value of δ can be taken as some fraction of r. The choice is somewhat arbitrary, but the value of 0.1 r seems reasonable; and the resulting value of α/δ allows one to calculate a value of D_A for the omega-mer which is probably correct to within a factor of two.[75] Thus, one obtains an estimate of α/δ equal to 2×10^3. Taking $r = 10^{-2}$ cm, fe = 0.4, $S = 0.5$ cm^2, and using the values of $\bar{V}_e$ and V_b given in Table X one obtains the values of R_{opt} shown in Table X . These values indicate that the optimum flowrate is at least three orders of magnitude below the value of R given in Table X. The practical use of such low flowrates is out of the question. In most cases the effect of lateral diffusion should not be a serious consideration. One can improve resolution by decreasing the flowrate to the limit of one's patience. Thus, the limiting resolution which one can obtain with a chromatographic system will usually be set by practical considerations, rather than the inherent physical limits of the technique.

In Chapter II.3 an empirical relation was given for the dependence of the elution position of an oligonucleotide upon its charge. The relation for $\bar{V}_e$ in Equation (20) gives a theoretical expression for this dependence provided one can find some way of relating K_A to the charge. In order to relate K_A to the charge, one re-examines the set of reactions shown in (1). For a polyanion, such as an

oligonucleotide, in which each charge is associated with a monomeric unit of the same basic molecular structure, one would expect the binding constants for all of the reactions to be very nearly the same. The possible exception would be the constant for the first reaction. One might expect the first constant to differ both because its units are slightly different and because its free energy would contain an additional entropy term. By allowing the first constant to differ from the remaining constants one can also take into account the fact that the terminal phosphate group possesses two negative charges. This should not be taken to imply that the terminal phosphate is the first to react. On the basis of the above observation one introduces the following simplified model. The binding constant for the initial reaction will be denoted by k_1. The binding constants for all succeeding reactions will be taken to be identical and will be denoted by k. From Equation (4) one sees that K_A will now be given by

$$K_A = k_1 k^{n-1} \sum_{r=0}^{n-1} \left(\frac{c_B}{kQ}\right)^r . \qquad (33)$$

Empirically one knows that a chromatographic separation will only be practical when the value of c_B is such that

$\rho K_A (Q/c_B)^n$ is larger than one. This fact is also predicted by the theory which has been derived (see Equation 30)). Under these conditions the sum in (33) can be approximated by its first term. This is equivalent to stating that, once a molecule becomes bound to the resin, all of its charged groups react with the resin. Even if c_B/kQ is not negligible compared to one, it will be much less dependent upon the elution volume than c_B^n. Thus, unless c_B/kQ is quite close to one, or larger, the approximation

$$k_A = k_1 k^{n-1} \tag{34}$$

should still be satisfactory although the physical significance of k_1 and k will be altered. Equation (20) can be rewritten as

$$\log \frac{\bar{V}_e}{V_b} \left(\frac{1}{n+1}\right)^{1/n+1} = - \frac{1}{n+1} \log\left(\frac{Qk}{gV_b} \frac{k}{\rho k_1}\right) + \log \frac{Qk}{gV_b} \tag{35}$$

A plot of the quantity on the left of Equation (35) against 1/n+1 should give a straight line. The slope and intercept can be used to evaluate k_1 and k. Figure 28 shows such a plot. The plot is linear for chainlengths up to nine at which point the slope becomes increasingly negative. The values of k_1 and k calculated from the slope and intercept of the plot are shown in Table X. However,

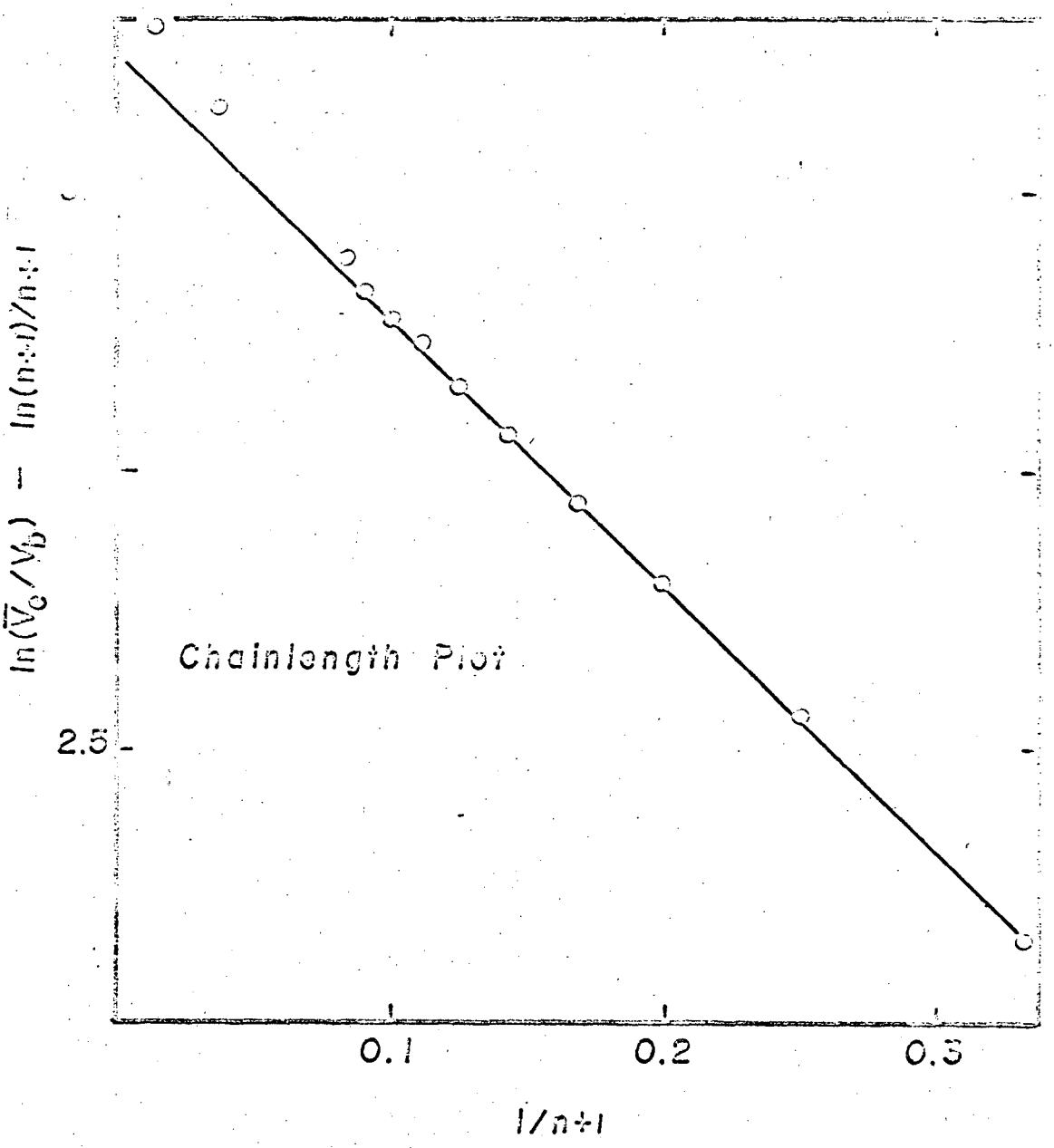


Figure 23. A test of the theoretical relationship between the charge of a species and its elution position in ion-exchange chromatography.

The experimental data is from the chainlength separation procedure for oligonucleotides described in Chapter II.3.

because of the direction of the deviation of the plot from linear behavior, one is uncertain about the physical significance of these constants.

The fact that the linearity breaks down is not surprising since the derivation of Equation (35) assumed the resin capacity, Q, to be independent of chainlength. The resin used for the separation was DEAE-Sephadex A-25 in which many of the ion-exchange groups are located in pores inside the resin particles. As the chainlength of the sample species increases, an increasingly smaller number of pores will be accessible to the species; and a decrease in the effective resin capacity will occur. However, the fact that the slope increases in magnitude is surprising since a decrease in available resin capacity should decrease the magnitude of the slope. The observed departure from linearity is almost certainly due to the fact that the approximation (34) becomes poorer as the chainlength increases. Using the value of k calculated from the intercept of the plot, one can calculate values for c_B/kQ at different chainlengths. The values begin at 0.28 for the monomer and increase until a value of 1.07 is reached at chainlength 70. Introducing a better approximation for the series in (33) yields a more complicated expression which can not be fit by a simple graphical method.

APPENDIX II

A Method of Linear Analysis for Arrays of Data -- with Emphasis on Collections of ORD and CD Spectra

It is the purpose of this section to outline a method by which the large amounts of data available from optical studies of polynucleotides can be effectively analyzed. Suppose that one has obtained a set of ORD spectra for a polymer at a number of temperatures. The proposed method of analysis allows one to decide how many significantly different shapes are contained in the set of ORD curves. Alternatively, one might say that it tells one the minimum number of component curves needed to fit all of the observed spectra to within experimental error. Since the method of fitting the experimental curves involves minimizing a sum of squares, it is natural to wonder how this method differs from ordinary methods of least squares analysis. The difference lies in the choice of the set of basis curves. One could, for example, fit the experimental spectra with a power series. However, the number of terms needed to fit the data to within experimental error would, in general, be larger than the number of linearly independent components contained in the set of spectra. The basis set used in this method is constructed from the set of experimental curves. This basis set results in a particularly elegant set of relations, and in a minimum number of components.

The utility of the information resulting from such a method of analysis is based on the following assumption. One assumes that an ORD or CD spectrum arising out of a specific geometrical conformation in a polymer will have a shape which is invariant with respect to changes in temperature and/or solvent composition. For example, suppose a polymer changes its conformation from form A to form B as one varies the ionic strength of the solvent from a to b. One assumes that the shape of the ORD spectrum associated with form A at an ionic strength intermediate between a and b, will be the same as at ionic strength a; and likewise for from B. No restriction is placed upon the magnitude of the spectra. The consequence of this assumption is that the ORD spectrum of the polymer at any ionic strength between a and b can be constructed by forming a linear combination of the spectra associated with the A and B forms.

Variations in the shape of a spectrum, such as red or blue shifts and the sharpening or broadening of a band, are usually small when one is dealing with simple molecules provided that the variation in temperature or solvent composition is over a fairly small interval (i.e., 0°C to 100°C for temperature or ionic strengths below 0.1). Thus, one can be fairly confident about assuming constant shape for polymer systems which can be described by a simple two-state model such as the example above. However, the constant shape assumption is also a valid approximation

for certain conformation changes of a multi-state nature. The torsional oscillator model for dinucleoside phosphates treated by Glaubiger, Lloyd, and Tinoco[76] gives rise to an ORD spectrum whose magnitude changes with temperature, but whose shape is constant. The shape of the ORD spectrum for this model is due to the splitting of two degenerate electronic transitions.[77] This shape will remain fairly constant so long as the splitting is small compared to the band width associated with the transitions.

On the basis of the assumption discussed above, one can associate the number of linearly independent components contained in a set of polymer spectra with the number of distinguishable conformations present in the polymer. If a situation occurs in which two or more conformations give rise to spectra with identical shape; or if the shapes are different but the temperature and/or solvent dependencies of the magnitudes are identical, then these configurations will be indistinguishable by the method of analysis being discussed. Thus, it will usually not be possible to pick out individual conformations in the polymer. Instead, one will observe classes of conformations. For example, it is unlikely that one could distinguish among the large number of different types of nearest neighbor interactions present in a large polynucleotide. The similarity of the temperature dependencies of the ORD spectra associated with these interactions would cause them to appear as a single component.

However, the nearest neighbor interactions could easily be distinguished from cooperative effects which possess a markedly different temperature dependence. Further discussion of the properties of the method will be put off until its derivation has been presented. It should be pointed out that, although the data discussed will be obtained from ORD or CD measurements, the method to be set forth is applicable to data from any source.

Suppose that one has measured the CD of a polymer at m different wavelengths and n different temperatures. This set of data can be put in the form of an array, $\underset{\sim}{S}$, such that each row of the array, $\underset{\sim}{S}_i$, will be the CD spectrum obtained at the i^{th} temperature. Thus, $\underset{\sim}{S}$, will contain n rows and m columns. A single row of the array, $\underset{\sim}{S}_i$, can be thought of as a vector in a m-dimensional space. The direction of the vector can be identified with the shape of the spectrum; and the length of the vector is a measure of the magnitude of the spectrum. One wishes to investigate the hypothesis that all of the experimental spectra can be expressed, to within the experimental error, by linear combinations of a set of μ basis spectra, where μ is a number less than n. If for each experimental spectrum one constructs the best possible linear combination of a given set of basis spectra, by the method of least squares, the sum of the squares of the deviations of each constructed spectrum from the corresponding experimental spectra, at

all wavelengths, will be equal to the number of degrees of freedom times the estimated variance of the experimental data.[78] The preceding statement contains the implied assumption that the error is normally distributed with a constant variance. A normal error curve is usually a satisfactory approximation,[78] and any change in the noise level with temperature or wavelength can be compensated by multiplying each experimental value by an appropriate weighting factor.

Let us now examine the situation outlined in the preceding paragraph from a geometrical standpoint. For a given $\underset{\sim}{S}_i$ and a given set of basis vectors, the sum of the squares of the deviations of $\underset{\sim}{S}_i$ from that linear combination which minimizes the sum of squares is by definition the norm (or square of the magnitude) of that component of $\underset{\sim}{S}_i$ which is orthogonal to all of the basis vectors. Thus, the estimated variance times the number of degrees of freedom will be given by the sum of the norms of the components, of the set of $\underset{\sim}{S}_i$, which are orthogonal to all of the basis vectors. Once one has obtained a set of basis vectors, one can perform the least squares fit without difficulty. However, one would like to find a basis set which contains the minimum number of components necessary to reduce the estimated variance to the level of the experimental error. It will be shown that this task can be accomplished in a very simple manner, and that in the process one can also obtain the coefficients needed for the least squares fit.

Consider the n by n symmetric matrix, $\underset{\sim}{S}\cdot\underset{\sim}{S}^{T}$, where T indicates the transpose of a matrix. This matrix will possess a set of n eigenvalues, e_i, and a corresponding set of n eigenvectors, $\underset{\sim}{U}_i$, whose components will be denoted by u_{ij}. One can prove the following three statements:

a) The best least squares fit of the set of $\underset{\sim}{S}_i$ with a set of μ basis vectors will be obtained when the basis vectors are given by

$$e_\alpha^{-\frac{1}{2}} \sum_{1}^{n} U_{\alpha i}\underset{\sim}{S}_i \equiv e_\alpha^{-\frac{1}{2}}\, \underset{\sim}{U}_\alpha\cdot\underset{\sim}{S}$$

provided that the set of e_α consists of the μ largest eigenvalues of $\underset{\sim}{S}\cdot\underset{\sim}{S}^{T}$.

b) The sum of the norms of the components, of the set of $\underset{\sim}{S}_i$, which are orthogonal to the space defined by the basis vectors in statement (a) will be given by the sum of the $n-\mu$ eigenvalues which were <u>not</u> used in constructing the basis vectors.

c) The coefficients for a least squares fit of a vector $\underset{\sim}{S}_k$ with a linear combination of the basis vectors in statement (a) will be given by $e_\alpha^{\frac{1}{2}}\, u_{\alpha k}$.

The above statements tell one how to construct the minimum set of basis vectors which will fit the experimental data to any desired degree of accuracy. All of the information needed for this process is contained in the eigenvalues and eigenvectors of the matrix $\underset{\sim}{S}\cdot\underset{\sim}{S}^{T}$. Thus, in practice,

all that is required is the construction of the matrix $\underset{\sim}{S}\cdot\underset{\sim}{S}^{T}$ from the experimental spectra and a subsequent diagonalization of this matrix. These operations can be performed with the Fortran program given at the end of this appendix. The practical use of this fitting procedure will be discussed further after the proof of the above statements has been demonstrated.

It is obvious that in constructing a set of basis vectors it is only necessary to consider those vectors which lie in the space defined by the experimental vectors. Thus, one need only consider sets of basis vectors of the form,

$$\sum_{i}^{n} x_{\alpha i}\,\underset{\sim}{S}_{i} \equiv \underset{\sim}{X}_{\alpha}\cdot\underset{\sim}{S}\ ; \qquad \alpha = 1,\cdots \mu \tag{1}$$

where $\underset{\sim}{X} = [x_{\alpha i}]$ is an n by μ matrix of, as yet, arbitrary coefficients, and $\underset{\sim}{X}_{\alpha}$ is the α^{th} row of this matrix. It is also sufficient to consider only those sets of basis vectors which are orthonormal. The orthonormality requirement places the following constraint on the coefficients $x_{\alpha i}$:

$$\left(\sum_{i}^{n} x_{\alpha i}\,\underset{\sim}{S}_{i}\right)\cdot\left(\sum_{j}^{m} x_{\beta j}\,\underset{\sim}{S}_{j}\right) \equiv \underset{\sim}{X}_{\alpha}\cdot\underset{\sim}{S}\cdot\underset{\sim}{S}^{T}\cdot X_{\beta}^{T} = \delta_{\alpha\beta} \tag{2}$$

The best set of μ basis vectors will be the one which minimizes the sum of the norms of the components, of the $\underset{\sim}{S}_{i}$, which are orthogonal to the set of basis vectors. The component, of a vector $\underset{\sim}{S}_{i}$, which is orthogonal to the basis set will be

$$\underset{\sim}{S}_i - \sum_{\alpha}^{\mu}\left(\underset{\sim}{S}_i \cdot \sum_{j}^{n} x_{\alpha j}\underset{\sim}{S}_j\right)\sum_{j}^{n} x_{\alpha j}\underset{\sim}{S}_j \equiv \underset{\sim}{S}_i - \sum_{\alpha}^{\mu}\left(\underset{\sim}{X}_\alpha \cdot \underset{\sim}{S} \cdot \underset{\sim}{S}_i^T\right)\underset{\sim}{X}_\alpha \cdot \underset{\sim}{S} \tag{3}$$

Using the relation in (2) one can show that the norm of this component will be given by

$$\underset{\sim}{S}_i \cdot \underset{\sim}{S}_i^T - \sum_{\alpha}^{\mu}(\underset{\sim}{X}_\alpha \cdot \underset{\sim}{S} \cdot \underset{\sim}{S}_i^T)(\underset{\sim}{S}_i \cdot \underset{\sim}{S}^T \cdot \underset{\sim}{X}_\alpha^T) \tag{4}$$

$$\text{where} \quad \underset{\sim}{X}_\alpha \cdot \underset{\sim}{S} \cdot \underset{\sim}{S}_i^T \equiv \underset{\sim}{S}_i \cdot \underset{\sim}{S}^T \cdot \underset{\sim}{X}_\alpha^T$$

Thus, the sum of the norms can be written as

$$\sum_{i}^{n}\left(\underset{\sim}{S}_i \cdot \underset{\sim}{S}_i^T - \sum_{\alpha}^{\mu}\underset{\sim}{X}_\alpha \cdot \underset{\sim}{S} \cdot \underset{\sim}{S}_i^T \underset{\sim}{S}_i \cdot \underset{\sim}{S}^T \cdot \underset{\sim}{X}_\alpha^T\right) \equiv T_r(\underset{\sim}{S} \cdot \underset{\sim}{S}^T) - \sum_{\alpha}^{\mu}\underset{\sim}{X}_\alpha \cdot (\underset{\sim}{S} \cdot \underset{\sim}{S}^T)^2 \cdot \underset{\sim}{X}_\alpha^T \tag{5}$$

This function must be minimized by varying the coefficients $x_{\alpha i}$ subject to the constraints in (2). The minimization is performed by the method of LaGrange.[79] One sets the partial derivatives, with respect to the set of $x_{\alpha i}$, of expressions (2) and (5) equal to zero, and introduces the LaGrange multipliers, $w_{\alpha\beta}$. The result is the equation

$$\underset{\sim}{X}_\alpha \cdot (\underset{\sim}{S} \cdot \underset{\sim}{S}^T)^2 = \sum_{\beta}^{\mu} w_{\alpha\beta}\underset{\sim}{X}_\beta \cdot \underset{\sim}{S} \cdot \underset{\sim}{S}^T \tag{6}$$

The LaGrange multipliers can be evaluated by forming the inner product between (6) and $\underset{\sim}{X}_\gamma^T$. One finds that

$$w_{\alpha\gamma} = \underset{\sim}{X}_\alpha \cdot (\underset{\sim}{S} \cdot \underset{\sim}{S}^T)^2 \cdot \underset{\sim}{X}_\gamma^T \tag{7}$$

Thus, the desired set of $x_{\alpha i}$ will be a solution to the matrix equation,

$$\underset{\sim}{X}_\alpha \cdot (\underset{\sim}{S} \cdot \underset{\sim}{S}^T)^2 = \sum_{\beta}^{\mu} (\underset{\sim}{X}_\alpha \cdot (\underset{\sim}{S} \cdot \underset{\sim}{S}^T)^2 \cdot \underset{\sim}{X}_\beta^T)\ \underset{\sim}{X}_\beta \cdot \underset{\sim}{S} \cdot \underset{\sim}{S}^T \tag{8}$$

One can show by substitution that equation (8) will be satisfied by any solution of the form

$$\underset{\sim}{X}_\alpha = \theta_\alpha^{-\frac{1}{2}}\ \underset{\sim}{U}_\alpha \tag{9}$$

where the set of $\underset{\sim}{U}_\alpha$ consists of any μ eigenvectors of $\underset{\sim}{S} \cdot \underset{\sim}{S}^T$. It is also true that any set of μ linear combinations, of μ eigenvectors, which satisfies equation (2) will be a solution. This merely reflects the fact that there are an infinite number of sets of orthonormal vectors which one can use to define a single space. Only those basis vectors formed from different sets of eigenvectors are distinct in the sense that they define different subspaces of the space defined by the set of $\underset{\sim}{S}_i$. There are $r!/\mu!(r-\mu)!$ such distince basis sets, where r is the rank of $\underset{\sim}{S} \cdot \underset{\sim}{S}^T$. It may be that there are other distinct basis sets. However, this question is unimportant since one can show that expression (5) possesses an absolute minimum when the set of θ_α in equation (9) contains the largest eigenvalues of $\underset{\sim}{S} \cdot \underset{\sim}{S}^T$.

Any set of $\underset{\sim}{X}_\alpha$ which is a solution to (8) can be written as a linear combination of the eigenvectors, $\underset{\sim}{U}_i$, as follows:

$$\underset{\sim}{X}_\alpha = \sum_{i}^{n} c_{\alpha i}\ \theta_i^{-\frac{1}{2}}\ \underset{\sim}{U}_i\ , \qquad \alpha = 1, \cdots, \mu \tag{10}$$

The value of expression (5) depends only upon the space defined by the set of basis vectors, not upon the particular

set of basis vectors used to define this space. The matrix of coefficients $c_{\alpha i}$ can always be put into the form,

$$\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ & c_{22} & \cdots & c_{2n} \\ & & \cdots & \\ & & \cdots & \\ & & & c_{un} \end{bmatrix} \tag{11}$$

by a unitary transformation acting only in the space of the basis vectors. Such a transformation changes only the representation of the space while leaving the space itself and expression (5) unchanged. Thus, one can choose a representation such that the eigenvalues of $\underset{\sim}{S}\cdot\underset{\sim}{S}^{T}$ will be indexed in order of descending magnitude, and $c_{\alpha i}$ will be equal to zero when i is less than α. Since the set of solutions, $\underset{\sim}{X}_{\alpha}$, must obey equation (2), one obtains the constraint,

$$\sum_{i}^{n} c_{\alpha i}\, c_{\beta i} = \delta_{\alpha\beta} \tag{12}$$

The above properties of the coefficients lead to the inequality,

$$\sum_{i}^{n} c_{\alpha i}^{2} e_{i} = \sum_{i=\alpha}^{n} c_{\alpha i}^{2} e_{i} \leq e_{\alpha} \sum_{i}^{n} c_{\alpha i}^{2} = e_{\alpha} \tag{13}$$

By substituting (10) into (5) and making use of (13), one finds that for any solution

$$T_r(\underset{\sim}{S}\cdot\underset{\sim}{S}^T) - \sum_{\alpha}^{\mu} \underset{\sim}{X}_\alpha \cdot (\underset{\sim}{S}\cdot\underset{\sim}{S}^T)^2 \cdot \underset{\sim}{X}_\alpha^T = \tag{14}$$

$$T_r(\underset{\sim}{S}\cdot\underset{\sim}{S}^T) - \sum_{\alpha}^{\mu}\sum_{i}^{n} c_{\alpha i}^2 e_i \geq T_r(\underset{\sim}{S}\cdot\underset{\sim}{S}^T) - \sum_{\alpha}^{\mu} e_\alpha =$$

$$\sum_{i=\mu+1}^{n} e_i$$

which completes the proof of statements (a) and (b). The proof of (c) is now trivial. For a set of orthogonal basis vectors the coefficients needed to fit a given $\underset{\sim}{S}_k$ with a linear combination of the basis vectors will just be the magnitudes of the projections of $\underset{\sim}{S}_k$ onto the basis vectors. Since the basis vectors are also normalized to unity, these projections will be given by the inner products of $\underset{\sim}{S}_k$ with the set of basis vectors,

$$\underset{\sim}{S}_k \cdot \sum_{i}^{n} e_\alpha^{-\frac{1}{2}} u_{\alpha i} \underset{\sim}{S}_i = e_\alpha^{\frac{1}{2}} U_{\alpha k} \tag{15}$$

One can now resume the discussion of the practical use of this method of analyzing sets of spectra. Suppose one has processed a set of data with the program at the end of this appendix, and obtained the eigenvectors and eigenvalues of $\underset{\sim}{S}\cdot\underset{\sim}{S}^T$. The first step is the determination of the minimum number of basis spectra needed to reduce the estimated variance to the level of the random noise in the measured spectra. One can estimate the variance provided that one knows the number of degrees of freedom involved in the fitting procedure. If one assumes that the error

in the initial data was not correlated in any way, each value of the CD will be an independent measurement, and the set of data will possess mr degrees of freedom. The value of r (the rank of $\underset{\sim}{S}\cdot\underset{\sim}{S}^T$) will be equal to n unless the initial set of spectra are not linearly independent, in which case r will be the number of non-zero eigenvalues of $\underset{\sim}{S}\cdot\underset{\sim}{S}^T$. If one now fits the data with μ vectors, each having m components, one will remove μm degrees of freedom. The number of remaining degrees of freedom will be $m(r-\mu)$, and the estimated variance for μ basis spectra will be

$$\sigma_\mu^2 = \frac{1}{m(r-\mu)} \sum_{i=\mu+1}^{n} e_i \,, \tag{16}$$

where the set of e_i are in descending order.

At this point one could merely guess the amount of noise in the data by inspection, calculate σ_μ^2 for $\mu=1,\cdots,r$ and choose the smallest value of μ for which σ_μ^2 was less than the value determined by inspection. However, there is a better method. The F-test[78] is a statistical test for the purpose of determining whether the decrease in the estimated variance caused by fitting a set of data to a more complicated model is significant. The test is applied ~~in the following manner~~. With μ basis vectors, one obtains an estimate of the variance, σ_μ^2, which is associated with a number of degrees of freedom, $m(r-\mu)$. Likewise, with $\mu+1$ components one estimates a variance, $\sigma_{\mu+1}^2$, less than σ_μ^2, associated with $m(r-\mu-1)$ degrees of freedom. Assuming

that the error of the experimental measurements is normally distributed, one can calculate[78] the probability, $P(F, f_1, f_2)$, that the ratio $\sigma_\mu^2/\sigma_{\mu+1}^2$ will be greater than or equal to some value F,

$$f_1 = m(r-\mu), \quad f_2 = m(r-\mu-1)$$

$$P(F,f_1,f_2) = \int_F^\infty \frac{\Gamma\left(\frac{f_1+f_2}{2}\right)\left(\frac{f_1}{f_2}F'\right)^{\frac{f_1}{2}-1}}{\Gamma\left(\frac{f_1}{2}\right)\Gamma\left(\frac{f_2}{2}\right)\left(1+\frac{f_1}{f_2}F'\right)^{\frac{f_1+f_2}{2}}} \frac{f_1}{f_2}\, dF' \qquad (17)$$

One usually finds values of F tabulated as a function of f_1 and f_2 for fixed values of $P(F, f_1, f_2)$. One might decide that only events with a probability below 0.01 represent a significant departure from random behavior. Then, if the value of $\sigma_\mu^2/\sigma_{\mu+1}^2$ exceeds the tabulated value of F for $P(F, f_1, f_2) = 0.01$, one concludes that the $\mu+1^{st}$ component is significant. If the ratio is below the tabulated value of F one concludes that the $\mu+1^{st}$ component is indistinguishable from the random error of the measurement.

The calculation of the number of degrees of freedom possessed by a set of data was made on the assumption that the error in the data was not correlated. If this is not the case, one will not usually know the degree of correlation; and it will be impossible to calculate the number of degrees of freedom. Correlation of the error can occur if the response time of the instrument with which the data is measured is too slow, or if the data is subjected to a

smoothing routine. The most serious source of correlation is that of baseline shifts. Unless one can obtain a reasonably stable baseline, one will have no hope of obtaining reasonable results from the F-test.

After one has determined the value of μ, one can use the eigenvectors and eigenvalues to perform a type of smoothing operation on the experimental data. One can replace the measured set of vectors, $\underset{\sim}{S}_i$, with a new set of vectors, $\underset{\sim}{S}'_i$, which are linear combinations, of the basis vectors, constructed by the method of least squares. The form of the coefficients for these linear combinations was given in equation (15). Thus, one obtains

$$\underset{\sim}{S}'_k = \sum_{\alpha}^{\mu} \left(e_{\alpha}^{\frac{1}{2}} \, u_{\alpha k} \right) \left(e_{\alpha}^{-\frac{1}{2}} \sum_{i}^{n} u_{\alpha i} \, \underset{\sim}{S}_i \right) \equiv \sum_{\alpha}^{\mu} u_{\alpha k} \, \underset{\sim}{U} \cdot \underset{\sim}{S} \qquad (18)$$

This smoothing process differs from other methods in that it makes no assumptions about the shapes of the spectra which are being smoothed.

As an illustration of this method of curve fitting, consider the two component system of artificially constructed spectra,

$$\underset{\sim}{S}_i = (2i)\underset{\sim}{J}_0(\tfrac{1}{2}k) + (10-2i)\underset{\sim}{J}_1(\tfrac{1}{2}k) + N\underset{\sim}{\Delta}_i; \qquad \begin{matrix} k = 1, \cdots, 10 \\ i = 0, \cdots, 5 \end{matrix} \qquad (19)$$

The components of the vectors $\underset{\sim}{J}_0(\frac{1}{2}k)$ and $\underset{\sim}{J}_1(\frac{1}{2}k)$, are the values of the Bessel functions of orders zero and one, respectively, for the indicated values of the argument. The components of the set of $\underset{\sim}{\Delta}_i$ are values taken from a table of normally distributed deviations,[80] and N is a

parameter used to vary the signal to noise ratio. This set of constructed data was analyzed at three different signal to noise ratios. The results of the F-test for each of the three cases are shown in Table XI. In each case the results indicate a value of μ equal to two. With a set of experimentally measured spectra, the results will usually not be quite as good. In practice, experimental measurements will usually contain some degree of systematic error in the baseline which will cause one's estimate of the number of degrees of freedom for the system to be incorrect.

For the purposes of comparison, Table XII gives the results obtained for an experimental two component system in which the magnitude of the baseline shift was approximately the same as the magnitude of the random noise. This system consisted of the CD spectra of six mixtures of poly A and poly C which one would not expect to interact at optical concentrations. The ratio of the estimated variances always remains higher than one would expect for the calculated number of degrees of freedom. Studies on other polymer systems, which are still in progress, indicate that one can usually obtain variance ratios which closely approach those of the ideal system discussed in the preceding paragraph.

One can use relation (15), together with the eigenvectors and eigenvalues shown in Table XIII, to calculate the magnitudes of the components of the $\underset{\sim}{S}_i$, in expression (19), along the two significant basis vectors. These values are related by the equation,

Table XI

Determination of the Number of Significant Basis Spectra for the System (J_0, J_1); m = 20, n = 6

μ	e_μ	degrees of freedom	σ_μ^2	$\sigma_\mu^2/\sigma_{\mu+1}^2$	F(5%)*
signal : noise = 3 ; $\sigma_\infty^2 = 1.0$					
0	960	120	10.32	3.72	1.39
1	190	100	2.78	2.54	1.42
2	37.2	80	1.09	1.30	1.50
3	19.6	60	0.838	1.09	1.68
4	18.9	40	0.766	1.31	1.99
5	11.7	20	0.586		
signal : noise = 10 ; $\sigma_\infty^2 = 0.09$					
0	961	120	9.58	5.11	1.39
1	179	100	1.88	18.51	1.42
2	3.37	80	0.101	1.28	1.50
3	1.98	60	0.0789	1.15	1.68
4	1.72	40	0.0689	1.32	1.99
5	1.04	20	0.0520		
signal : noise = 30 ; $\sigma_\infty^2 = 0.01$					
0	965	120	9.56	5.25	1.39
1	181	100	1.82	161.00	1.42
2	0.373	80	0.0113	1.27	1.50
3	0.228	60	0.00888	1.17	1.68
4	0.193	40	0.00762	1.36	1.99
5	0.112	20	0.00562		

* The probability that the ratio of variances will exceed this value is 5%.

Table XII

Determination of the Number of Significant Basis Spectra for a Collection of ORD Spectra of Poly A and Poly C at 25°C. in 0.005 M NaCl, 0.001 M Na-Cacodylate, pH 7.0; m = 72, n = 6

μ	e_μ	degrees of freedom	σ_μ^2	$\sigma_\mu^2/\sigma_{\mu+1}^2$	F(5%)*	F(1%)**
0	5820	432	14.7	9.75	1.18	1.29
1	532	360	1.51	35.0	1.22	1.33
2	8.51	288	0.0432	2.38	1.25	1.37
3	2.96	216	0.0182	2.72	1.29	1.43
4	0.822	144	0.00667	3.49	1.42	1.65
5	0.138	72	0.00191			

*The probability that the ratio of variances will exceed this value is 5%.

**The probability that the ratio of variances will exceed this value is 1%.

$$\begin{bmatrix} u_{1i} \\ u_{2i} \end{bmatrix} = \begin{bmatrix} u_{10} & u_{15} \\ u_{20} & u_{25} \end{bmatrix} \cdot \begin{bmatrix} c_{0i} \\ c_{5i} \end{bmatrix} \qquad (20)$$

to the coefficients, c_{0i} and c_{5i}. The coefficient, c_{0i}, is the magnitude of the component of $\underset{\sim}{S}_i$ along $\underset{\sim}{S}'_0$ divided by the magnitude of $\underset{\sim}{S}'_0$, where the prime denotes a vector which has been smoothed according to equation (18). The coefficient, c_{5i} is defined in a similar manner. The calculated values are compared, in Table XIII, with the theoretical values (those which one would obtain in the absence of the error term, $\underset{\sim}{\Delta}_i$).

Table XIII also includes the values of the analogous coefficients which one obtains by performing a least squares fit of the set of $\underset{\sim}{S}_i$ to the unsmoothed spectra $\underset{\sim}{S}_0$ and $\underset{\sim}{S}_5$. It is the smoothing process which differentiates the eigenvector approach from other matrix treatments of experimental data, such as matrix rank analysis.[81] If a set of spectra are free of experimental error, one can try to fit n-1 of the spectra with one spectrum. If this does not work, one can try fitting with two spectra. This process can be continued until one obtains a set of μ spectra which can be used to fit the n-μ remaining spectra. Matrix rank analysis is a systematic method of performing the above process. This technique has been used by McMullen, Jaskunas, and Tinoco[32] to analyze the variation in the ORD of TMV-RNA

Table XIII

The Fit of the Two Component System (J_0, J_1) to the Basis Vectors (U_1, U_2)

i	u_{1i}	u_{2i}	$c_{0,i}$			$c_{5,i}$		
			Theor.	eigen-vectors	least squares	Theor.	eigen-vectors	least squares
signal : noise = 3								
0	.328	-.555	1.000	1.000	1.000	0.000	0.000	0.000
1	.320	-.380	0.800	0.659	0.628	0.200	0.324	0.325
2	.399	-.373	0.600	01564	0.549	0.400	0.351	0.318
3	.407	.137	0.400	0.307	0.225	0.600	0.791	0.784
4	.446	.208	0.200	0.111	0.161	0.800	0.801	0.662
5	.516	.589	0.000	0.000	0.000	1.000	1.000	1.000
signal : noise = 10								
0	.339	-.624	1.000	1.000	1.000	0.000	0.000	0.000
1	.354	-.399	0.800	0.758	0.753	0.200	0.229	0.232
2	.395	-.214	0.600	0.594	0.592	0.400	0.382	0.378
3	.416	.096	0.400	0.353	0.352	0.600	0.661	0.655
4	.444	.283	0.200	0.174	0.179	0.800	0.796	0.783
5	.483	.562	0.000	0.000	0.000	1.000	1.000	1.000
signal : noise = 30								
0	.343	-.633	1.000	1.000	1.000	0.000	0.000	0.000
1	.364	-.398	0.800	0.786	0.785	0.200	0.209	0.210
2	.394	-.178	0.600	0.598	0.598	0.400	0.394	0.394
3	.418	.083	0.400	0.385	0.385	0.600	0.618	0.618
4	.443	.304	0.200	0.192	0.192	0.800	0.797	0.796
5	.473	.556	0.000	0.000	0.000	1.000	1.000	1.000

as a function of temperature and ionic strength. So long as the effects for which one is looking are large compared to the experimental error, such a method is entirely satisfactory. However, matrix rank analysis possesses a defect which becomes serious when one tries to look at components which are only an order of magnitude above the error level. This defect arises because preference is given to those spectra which are used to do the fitting. When one chooses a particular spectrum to fit the remaining spectra, this curve is treated as though it contained no error. If it does contain a certain level of error, this error is transferred to the remaining spectra. Thus, as one proceeds with the analysis, the error is propagated. The method of analysis described here avoids this problem by giving equal weight to every spectrum in the set.

The following program performs the diagonalization necessary to generate the eigenvectors and eigenvalues used in the preceding appendix. The program is written in Chippewa Fortran for use with the Control Data 6600 Computer.

```
      PROGRAM FTEST (INPUT,OUTPUT)
C     THIS IS PROGRAM ONE, IT COMPUTES THE EIGEN VALUES FOR
C     THE LEAST SQUARES FIT AND PERFORMS CHECKS
C     ON THE DIAGONALIZATION ROUTINE
C
      DIMENSION ORD(6,84), A(6,6), U(6,6)
      COMMON N,II,A,U
  101 FORMAT (2I4)
  102 FORMAT (12F6.3)
  201 FORMAT(///* THIS IS THE ORIGINAL FORM OF A(IK)*//)
  202 FORMAT(///* THIS IS U(IJ)A(JJ)U(KJ), DOES IT EQUAL*/
     1* A(IK)*//)
  203 FORMAT(///* THIS IS U(IJ)U(KJ), DOES IT EQUAL THE UNIT*/
     1* MATRIX*//)
  204 FORMAT(///* THIS IS U(JI)U(JK), DOES IT EQUAL THE UNIT*/
     1* MATRIX*//)
  205 FORMAT(///* ALLCHECKS HAVE BEEN COMPLETED, THE OUTPUT*/
     1* DATA FOLLOWS *//* THIS IS U(IK), ROWS*/
     1* ARE IN THE ORDER IN WHICH SPECTRA WERE*/
     1* READ, COLUMNS ARE ORDERED THE SAME*/
     1* AS THE EIGEN VALUES*//)
  206 FORMAT(///*THIS IS THE LIST OF EIGEN VALUES*//)
  207 FORMAT (E20.8/)
      READ 101, M,N
      READ 102, ( (ORD(I,J), J=1,M), I=1,N)
      DO 2 I=1,N
      DO 2 K=1,N
      A(I,K) = 0
      DO 1 J=1,M
      A(I,K) = A(I,K) + ORD(I,J)*ORD(K,J)
    1 CONTINUE
    2 U(I,K) = A(I,K)
      PRINT 201
      II = 2
      CALL PRIMAT
      CALL JACVAT(0)
      DO 3 J=1,N
    3 ORD(J,1) = A(J,J)
      DO 4 I=1,N
      DO 4 K=1,N
      A(I,K) = 0
      DO 4 J=1,N
    4 A(I,K) = A(I,K) + U(I,J)*ORD(J,1)*U(K,J)
      PRINT 202
      II = 1
      CALL PRIMAT
      DO 5 I=1,N
      DO 5 K=1,N
      A(I,K) = 0
      DO 5 J=1,N
    5 A(I,K) = A(I,K) + U(I,J)*U(K,J)
      PRINT 203
      CALL PRIMAT
```

```
      DO 6 I=1,N
      DO 6 K=1,N
      A(I,K) = 0
      DO 6 J=1,N
    6 A(I,K) = A(I,K) + U(J,I)*U(J,K)
      PRINT 204
      CALL PRIMAT
      PRINT 205
      II = 2
      CALL PRIMAT
      PRINT 206
      PRINT 207, (ORD(J,1), J=1,N)
      STOP
      END
```

```
      SUBROUTINE PRIMAT
      COMMON N,II,A,U
      DIMENSION A(6,6),U(6,6)
  301 FORMAT  (6(18H          COLUMN   ,I2) / )
  304 FORMAT (6E20.8 /)
      IK = -5
    1 IK = IK+6
      IJ = IK + 5
      IF (N-IJ) 2,3,3
    2 IJ = N
    3 PRINT 301, (K, K=IK,IJ)
      DO 6 I=1,N
      GO TO (4,5), II
    4 PRINT 304, (A(I,K), K=IK,IJ)
      GO TO 6
    5 PRINT 304, (U(I,K), K=IK,IJ)
    6 CONTINUE
      IF (N-IJ) 7,7,1
    7 RETURN
      END
```

```
      SUBROUTINE JACVAT(IEGEN)
      COMMON N, II, A, EIVR
      DIMENSION A(6,6), EIVR(6,6)
      IF(N-1) 2,2,1
    2 EIVR(1,1)=1.0
      RETURN
    1 IF(IEGEN) 102,99,102
   99 DO 101 J=1,N
      DO 100 I=1,N
  100 EIVR(I,J)=0.0
  101 EIVR(J,J)=1.0

C         FIND THE ABSOLUTELY LARGEST ELEMENT OF A

  102 ATOP=0.
      DO 111 I=1,N
      DO 111J=I,N
      IF(ATOP-ABS(A(I,J)))104,111,111
  104 ATOP=ABS(A(I,J))
  111 CONTINUE
      IF(ATOP)109,109,113
  109 RETURN

C         CALCULATE THE STOPPING CRITERION -- DSTOP

  113 AVGF=FLOAT(N*(N-1))*.55
      D=0.0
      DO 114 JJ=2,N
      DO 114 II=2,JJ
      S=A(II-1,JJ)/ATOP
  114 D=S*S+D
      DSTOP=(1.E-06)*D

C         CALCULATE THE THRESHOLD, THRSH
C
      THRSH = SQRT(D/AVGF)*ATOP
C
C         START A SWEEP
C
  115 IFLAG=0
      DO 130 JCOL=2,N
      JCOL1=JCOL-1
      DO 130 IROW=1,JCOL1
      AIJ=A(IROW,JCOL)
C
C         COMPARE THE OFF-DIAGONAL ELEMENT WITH THRSH
C
      IF(ABS(AIJ)-THRSH)130,130,117
  117 AII=A(IROW,IROW)
      AJJ=A(JCOL,JCOL)
      S=AJJ-AII
C
C     CHECK TO SEE IF THE CHOSEN ROTATION IS LESS THAN
C     THE ROUNDING ERROR.
C
```

```
      IF(ABS(AIJ)-1.E-09*ABS(S))130,130,118
  118 IFLAG=1
C
C     IF THE ROTATION IS VERY CLOSE TO 45 DEGREES, SET
C     SIN AND COS TO 1/(ROOT 2).
C
      IF(1.E-10*ABS(AIJ)-ABS(S))116,119,119
  119 S=.70710678118655
      C=S
      GO TO 120
C
C     CALCULATION OF SIN AND COS FOR ROTATION THAT IS
C     NOT VERY CLOSE TO 45 DEGREES
C
  116 T=AIJ/S
      S=0.25/SQRT(0.25+T*T)
C
C        COS = C ,  SIN= S
C
      C=SQRT(0.5+S)
      S=2.*T*S/C
C
C        CALCULATION OF THE NEW ELEMENTS OF MATRIX A
C
  120 DO 121 I=1,IROW
      T=A(I,IROW)
      U=A(I,JCOL)
      A(I,IROW)=C*T-S*U
  121 A(I,JCOL)=S*T+C*U
      I2=IROW+2
      IF(I2-JCOL)127,127,123
  127 CONTINUE
      DO 122 I=I2,JCOL
      T=A(I-1,JCOL)
      U=A(IROW,I-1)
      A(I-1,JCOL)=S*U+C*T
  122 A(IROW,I-1)=C*U-S*T
  123 A(JCOL,JCOL)=S*AIJ+C*AJJ
      A(IROW,IROW)=C*A(IROW,IROW)-S*(C*AIJ-S*AJJ)
      DO 124 J=JCOL,N
      T=A(IROW,J)
      U=A(JCOL,J)
      A(IROW,J)=C*T-S*U
  124 A(JCOL,J)=S*T+C*U
C
C        ROTATION COMPLETED.
C        SEE IF EIGENVECTORS ARE WANTED BY USER
C
      IF(IEGEN) 126,131,126
  131 DO 125 I=1,N
      T=EIVR(I,IROW)
      EIVR(I,IROW)=C*T-EIVR(I,JCOL)*S
  125 EIVR(I,JCOL)=S*T+EIVR(I,JCOL)*C
```

```
C
C          CALCULATE THE NEW NORM D AND COMPARE WITH DSTOP
C
  126 CONTINUE
      S=AIJ/ATOP
      D=D-S*S
      IF(D-DSTOP)1260,129,129
C
C     RECALCULATE DSTOP AND THRSH TO DISCARD
C     ROUNDING ERRORS
C
 1260 D=0.
      DO 128 JJ=2,N
      DO 128 II=2,JJ
      S=A(II-1,JJ)/ATOP
  128 D=S*S+D
      DSTOP=(1.E-06)*D
  129 THRSH=SQRT(D/AVGF)*ATOP
  130 CONTINUE
      IF(IFLAG)115,134,115
  134 RETURN
      END
```

REFERENCES

1. Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R., and Zamir, A., Science, 147, 1462 (1965).
2. Raj Bhandary, U. L., Chang, S. H., Stuart, A., Faulkner, R. D., Hoskinson, R. M., and Khorana, H. G., Proc. Natl. Acad. Sci., 57, 751 (1967).
3. Dube, S. K., Marcher, K. A., Clark, B. F. C., and Cory, S., Nature, 218, 232 (1968).
4. Brownlee, G. G. and Sanger, F., J. Mol. Biol., 23, 337 (1967).
5. Hall, J. B., Sedat, J. W., Adiga, P. R., Uemura, I., and Winnick, T., J. Mol. Biol., 12, 162 (1965)
6. Sedat, J. W. and Hall, J. B., J. Mol. Biol., 12, 174 (1965).
7. Mandeles, S., J. Biol. Chem., 242, 3103 (1967).
8. Steinschneider, A. and Fraenkel-Conrat, H., Biochemistry, 5, 2735 (1966).
9. de Wachter, R. and Fiers, W., J. Mol. Biol., 30,507 (1967).
10. Mandeles, S., private communications.
11. Mundry, K. W., Z. Vererbungsl., 97, 281 (1965).
12. Mandeles, S., J. Biol. Chem., 243,3671 (1968).
13. Hiester, N. K., and Vermeulen, T., Chem. Eng. Prog., 48, 505 (1952).
14. Vermeulen, T. and Hiester, N. K., Ind. Eng. Chem., 44, 636 (1952).

15. Bartos, E. M., Rushizky, G. W., and Sober, H. A., Biochemistry, 2, 1179 (1963).

16. Holley, R. W., Apgar, J., and Merrill, S. H., J. Biol. Chem., 236, PC 42 (1961).

17. Neu, H. C. and Heppel, L. A., J. Biol. Chem., 239, 3893 (1964).

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

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

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

21. Warshaw, M. M. and Tinoco, I. Jr., J. Mol. Biol., 13, 54 (1965).

22. Cantor, C. R. and Tinoco, I. Jr., J. Mol. Biol., 13, 65 (1965).

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

24. Tennent, H. G. and Vilbrandt, C. F., J. Am. Chem. Soc., 65, 424 (1943).

25. Van Holde, K. E. and Baldwin, R. L., J. Phys. Chem., 62, 734 (1958).

26. Johnson, J. S., Kraus, K. A., and Scatchard, G., J. Phys. Chem., 58, 1034 (1954).

27. Inman, R. B. and Jordan, D. O., Biochim. Biophys. Acta, 42, 421 (1960).

28. Felsenfeld, G. and Hirschmann, G., J. Mol. Biol., 13, 407 (1965).

29. Fresco, J. L., Klotz, L. C., and Richards, E. G., Cold Spring Harbor Symp. Quant. Biol., 28, 83 (1963).

30. Bradley, D. F., Tinoco, I. Jr., and Woody, R. W., Biopolymers, 1, 239 (1963).

31. Tinoco, I. Jr., Woody, R. W., and Bradley, D. F., J. Chem. Phys., 38, 1317 (1963).

32. McMullen, D. W., Jaskunas, S. R., and Tinoco, I. Jr., Biopolymers, 5, 589 (1967).

33. Savitzky, A. and Golay, M. J. E., Anal. Chem., 36, 1627 (1964).

34. Fresco, J. R., Alberts, B. M., and Doty, P., Nature, 188, 98 (1960).

35. Fresco, J. R., Tetrahderon, 13, 185 (1961).

36. Holcomb, D. N. and Tinoco, I. Jr., Biopolymers, 3, 121 (1965).

37. McDonald, C. C. and Phillips, W. D., Science, 144, 1234 (1964).

38. Brahms, J. and Mommearts, W. J., J. Mol. Biol., 10, 73 (1964).

39. Davis, R. C. and Tinoco, I. Jr., Biopolymers, 6, 223 (1968).

40. McDonald, C. C., Phillips, W. D., and Lazar, J., J. Am. Chem. Soc., 89, 4166 (1967).

41. Cantor, C. R., Ph.D. thesis, University of California (1966).

42. Jaskunas, S. R., Ph.D. thesis, University of California (1968).

43. Egami, F., Takahashi, E., and Uchida, T., Prog. Nuc. Acid Res., 3, 82 (1964).

44. Volkin, E. and Cohn, W. E., J. Biol. Chem., 205, 767 (1953).

45. Anfinsen, C. B. and White, F. H., The Enzymes, Vol. 5, 2nd Ed., Boyer, P. D., Lardy, H., and Myrback, K., ed., Academic Press, New York (1961), p. 95.

46. Hakin, A. A., J. Biol. Chem., 228, 459 (1957).

47. Rushizky, G. W. and Knight, C. A., Virology, 11, 236 (1960).

48. Beers, R. F. Jr., J. Biol. Chem., 235, 2393 (1960).

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

50. Stanley, W., Jr., Ph.D. thesis, University of Wisconsin (1964).

51. Warshaw, M. M., Ph.D. thesis, University of California (1965).

52. Gilham, P. T., J. Am. Chem. Soc., 84, 687 (1962).

53. Lee, J. C., Ho, N. W. R., and Gilham, P. T., Biochim. Biophys. Acta, 95, 503 (1965).

54. Ho, N. W. Y. and Gilham, P. T., Biochemistry, 6, 3632 (1967).

55. Anderson, J. H. and Carter, C. E., Biochemistry, 4, 1102 (1965).

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

57. Hart, R. G., Proc. Natl. Acad. Sci., 41, 261 (1955).

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

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

60. Kado, C. I. and Knight, C. A., Biochemistry, 55, 1276 (1966).

61. Anderer, F. A. and Handschuh, D., Z. Naturforsch., 17b, 536 (1962).

62. Funatsu, G., Tsugita, A., and Fraenkel-Conrat, H., Arch. Biochem. Biophys., 105, 25 (1964).

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

64. Weissman, C., Billeter, M. A., Schneider, M. C., Knight, C. A., and Ochoa, S., Proc. Natl. Acad. Sci., 53, 653 (1965).

65. Anfinsen, C. G., The Molecular Basis of Evolution, John Wiley, New York (1961).

66. Smyth, D. G., Stein, W. H., and Moore, S., J. Biol. Chem., 238, 227 (1963).

67. Takahashi, K., J. Biol. Chem., 240, PC 4117 (1965).

68. Cantor, C. and Jukes, T., Biochem. Biophys. Res. Comm., 23, 319 (1966).

69. Mandeles, S. and Tinoco, I. Jr., Biopolymers, 1, 183 (1963).

70. Holley, R. W., Madison, J. T., and Zamir, A., Biochem. Biophys. Res. Comm., 17, 389 (1964).

71. Takanami, M., J. Mol. Biol., 23, 135 (1967).

72. Khorana, H. G., The Enzymes, Vol. 5, 2nd ed., Boyer, P. D., Lardy, H., and Myrback, K., ed., Academic Press, New York (1961), p. 79.

73. Thomas, H. C., J. Amer. Chem. Soc., 66, 1664 (1944).

74. Walter, J. E., J. Chem. Phys., 13, 332 (1945).

75. Tanford, C., Physical Chemistry of Macromolecules, John Wiley, New York (1967), p. 358.

76. Glaubiger, D., Lloyd, D. A., and Tinoco, I. Jr., Biopolymers, 6, 409 (1968).

77. Bush, C. A. and Tinoco, I. Jr., J. Mol. Biol., 23, 601 (1967).

78. Fraser, D. A. S., Statistics An Introduction, John Wiley, New York (1960), pp. 68, 202, 233.

79. Lass, H., Elements of Pure and Applied Mathematics, McGraw-Hill, New York (1957), pp. 296, 467.

80. Rand Corporation, A Million Random Digits with 100,000 Normal Deviates, Free Press, Glencoe, Ill. (1955).

81. Wallace, R. and Katz, S. M., J. Phys. Chem., 68, 3890 (1964).

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