UCRL-16701

University of California

Ernest O. Lawrence Radiation Laboratory

SEQUENCE DEPENDENT PROPERTIES OF OLIGONUCLEOTIDES

Berkeley, California

DISCLAIMER

UCRL-16701

UNIVERSITY OF CALIFORNIA

Lawrence Radiation Laboratory
Berkeley, California

AEC Contract No. W-7405-eng-48

SEQUENCE DEPENDENT PROPERTIES OF OLIGONUCLEOTIDES

Charles Robert Cantor

Ph. D. Thesis

January 1966

This thesis is dedicated to Siphonaptera Tunga tiggerans, whose attention has been of inestimable help.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

OLIGONUCLEOTIDES

A	Adenosine
U	Uridine
C	Cytidine
G	Guanosine
I	Inosine
T	Ribothymidine
ψ	Pseudouridine
DiHU	Dihydrouridine
N	A general nucleoside
Np	3'-nucleotide
pN	5'-nucleotide
N>p	2',3'-cyclic nucleotide
Nx	2'-O-methylnucleoside
NDP	Nucleoside diphosphate
ppN	Nucleoside diphosphate
NTP	Nucleoside triphosphate
NpN	Dinucleoside phosphate
dN	Deoxynucleoside

NUCLEIC ACIDS AND POLYNUCLEOTIDES

RNA	Ribonucleic acid
DNA	Deoxyribonucleic acid
mRNA	Messenger RNA
rRNA	Ribosomal RNA
sRNA	Amino-acyl transfer RNA
poly N	Homopolynucleotide
poly (A+U)	1:1 complex of poly A and poly U
poly AU	1:1 random copolymer of A and U
poly r(AU)	Polymer with regularly alternating sequence
poly d(AT)	Polymer with regularly alternating sequence

MISCELLANEOUS

RNAase	Ribonuclease
DNAase	Deoxyribonuclease
PDE	Phosphodiesterase
TMV	Tobacco Mosaic Virus
TYMV	Turnip Yellow Mosaic Virus
ala	Alanine
EDTA	Ethylenediaminetetraacetate
Tris	Tris(hydroxymethyl)aminomethane
ORD	Optical rotatory dispersion
CD	Circular dichroism
P_i	Inorganic phosphate

ABSTRACT

SEQUENCE DEPENDENT PROPERTIES OF OLIGONUCLEOTIDES

Charles Robert Cantor

In this work we discuss some of the ways in which the base sequence of an oligoribonucleotide can affect its physical and chemical properties. The base sequence can have direct effects on the absorption spectra and optical rotatory dispersion of single strand oligonucleotides. It will also manifest itself directly by determining the pattern of mononucleotides released when the oligonucleotide is exposed to an exonuclease. Since the conformation of an oligonucleotide can be very sensitive to base sequence, many additional properties of double strand oligonucleotides may be expected to depend strongly on the nucleotide sequence.

We explore the possibilities of using exonuclease degradation to determine base sequence. Calculations of the kinetics of two models of enzyme attack are described. It is shown that the sequence information can be most clearly obtained if the enzyme attack can proceed in a random fashion, and if the enzyme is present in excess. The optimal choice of parameters may permit the use of exoenzymes to determine the sequence of the 15-20 terminal residues of a polynucleotide chain.

We have measured the absorption spectra and optical rotatory dispersion of seven trinucleoside diphosphates. These results demonstrate that the latter property is quite sensitive to base sequence. We show that it is possible to calculate the optical properties of trinucleoside

diphosphates from the experimentally determined properties of their component dinucleoside phosphates and mononucleosides. This is done by making the assumptions that only nearest neighbor interactions are important, and that the conformation of neighboring bases is independent of chain length. When these calculations are extended to single strand polynucleotides the results are still in reasonable qualitative agreement with experiment.

The optical rotatory dispersion of TMV RNA in the absence of salt is in good agreement with the results calculated from dimer and monomer properties. Thus under these conditions a single stacked strand is probably a good description of the molecular conformation. It is possible to make an estimate of the changes in optical rotation which accompany the formation of base pairs in RNA. From these estimates, and the measured properties, we can show that, in the presence of salt, TMV RNA probably has A-U and G-C base pairs. In a similar way we can estimate the optical rotatory dispersion of various models which have been proposed for the conformation of the alanine sRNA. We discuss, in detail, the properties of one model which we have constructed which contains 26 base pairs. This model is in good agreement with most of the available experimental evidence for the conformation of the alanine sRNA.

I. INTRODUCTION

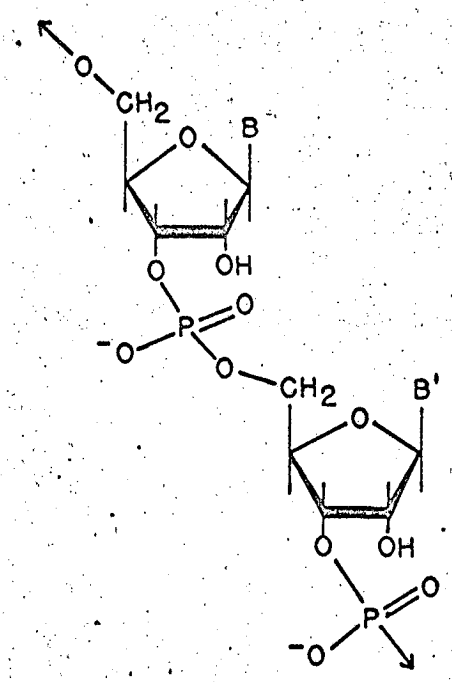
One of the most challenging problems for the physical chemist interested in biology is the elucidation of the structure and properties of the macromolecules which are found to play a critical role in the chemistry of living organisms. Two types of macromolecules which have received considerable attention in the last decade are deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The general biological role of these large polymers is well established, but their chemistry is not yet well enough understood to permit a detailed explanation of their mechanism of action. Thus our long range goal is to be able to attain a chemical and physical explanation for the biological properties of nucleic acids. In addition, we hope to be able to understand why these molecules are especially suited for their biological role. These are far reaching questions which cannot be answered by a single experiment, a single group of experimenters, or even a single discipline. But we hope that our work will prove to be helpful and of interest to those who share our goals.

The first step towards developing a chemical understanding of the biology of nucleic acids is the determination of their structure. What we would really like to know is the structure in vivo, but for the present we must be content with trying to determine the structure in aqueous solution. A further simplification would be to study the molecular structure in a crystal or fiber, but thus far this has proven feasible generally only for DNA. In the present context, "structure" means both the chemical structural formula and the conformation of the molecule in three dimensions. For a molecule as

large as RNA or DNA it is apparent that the conformation plays a critical role in determining the biological activity. Once the structure of nucleic acids has been determined in solution it should be possible to fruitfully apply many of the chemical and physical methods at our disposal to explain the biological properties of these macromolecules.

In this work we shall be concerned almost exclusively with the structure and properties of RNA. But many analogies can certainly be drawn between what we have found for RNA and what is known or expected in DNA. The primary structure of RNA has been known in general for quite some time. It consists of a phosphate D-ribose backbone with various bases attached to the 1' carbon of the sugar. This structure is illustrated in Figure 1 along with the commonly used schematic abbreviation. In this notation, B represents a nucleoside. Thus Bp stands for a 3' nucleotide and pB is a 5' nucleotide. The four most common nucleosides are adenosine (A), guanosine (G), uridine (U), and cytidine (C), and many RNA's contain only these four. The structures and abbreviations of these and other nucleosides we shall discuss are shown in Figure 2. The aspect of the primary structure which remains undetermined for most nucleic acids is the sequence of bases along the polynucleotide chain. The base sequence is known for only one RNA.[97] In Chapter II we shall discuss some of the methods of base sequence determination which have been used or proposed. We shall evaluate the use of exonuclease degradation in detail, and we show that under some conditions this can be a very profitable technique.

The second major problem of RNA structure is the determination of the conformation under a variety of experimental conditions which we hope approximate the <u>in vivo</u> state. A major complication is the wide



-----BpB'p-------

POLYRIBONUCLEOTIDE ABBREVIATION

MU-37120

Figure 1. The structure of polyribonucleotides.

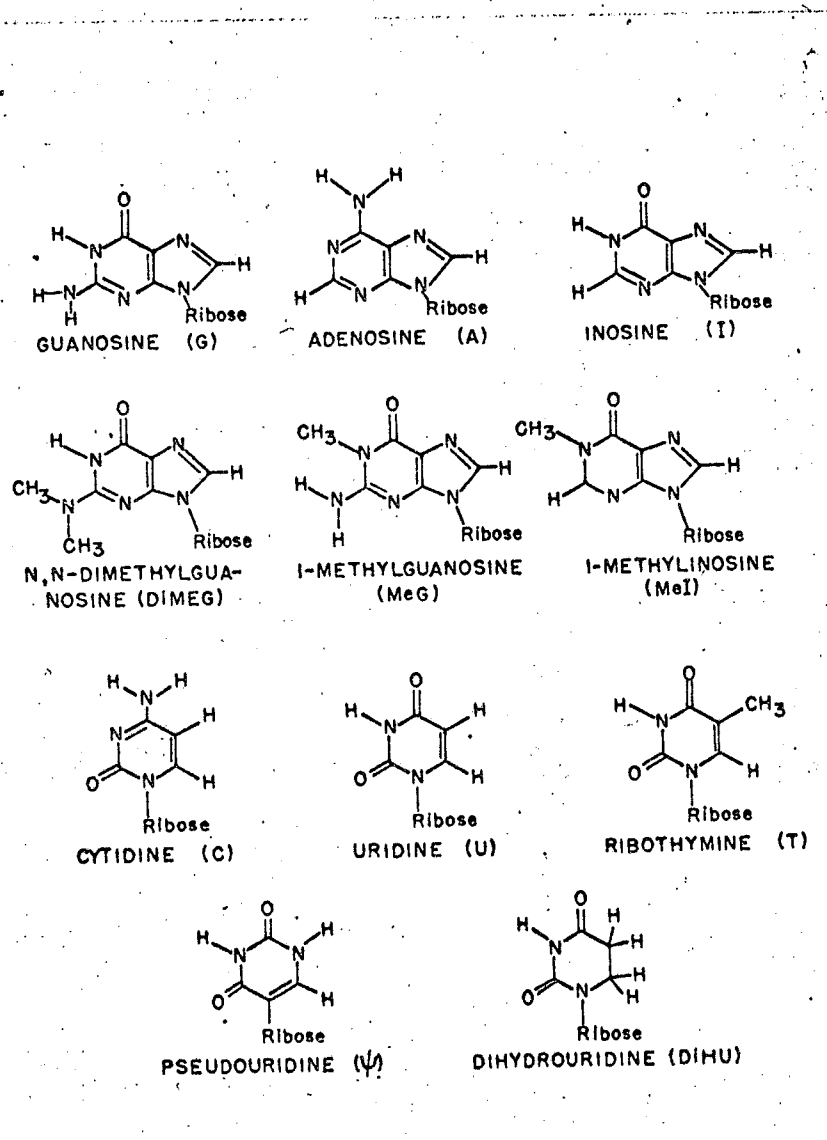


Figure 2. Structures of some of the nucleosides found in RNA.

range of sizes and functions known to exist for RNA. Messenger RNA (mRNA) is responsible for carrying the genetic information from the DNA in the nucleus to the protein synthesizing system located on the ribosome. A major component of the ribosome is ribosomal RNA (rRNA), which presumably contributes to the mechanism of protein synthesis. Soluble or amino acid transfer RNA (sRNA) is involved in transporting amino acids to the ribosome, and may help to affect the translation of the nucleic acid code into protein sequence. Thus the functions of RNA are vital to the cell. The range of molecular sizes found for natural RNA's varies from the gramicidin S mRNA with a molecular weight of 11,200[201] to the RNA from Reovirus which has a molecular weight of about 10^7.[118] The base composition of RNA can vary from 7S virus RNA, with almost random proportions of the four common bases (A:U:C:G = 238:258:256:246)[176] to human liver RNA, which has a great proponderance of G and C (A:U:C:G = 115:126:316:441).[172] Along with this great variation in primary structure, a large number of different conformations have been suggested for RNA. These include multiple strand helices, double strand helices, single strand helices, random coils and more complicated folded arrangements.[217] In Chapter IV, we shall show that studies of the optical rotatory dispersion (ORD) can be of great help in differentiating among these possibilities.

The conformation of an RNA in solution is a problem which should not be over-simplified. Overly ambitious or naive studies on RNA may lead to a frustrating situation in which countless possibilities cannot be distinguished experimentally, and the uncertainties of our intuition must be called upon to solve the dilemma. What is needed is a set of simple model systems. The results found should stimulate further ex-

periments with polymers, and, at the same time, may help us to refine our intuition. The simple models for RNA are the oligoribonucleotides. We expect that most of the properties of RNA should eventually be explainable in terms of the properties of its smaller components. The great advantage in studying oligonucleotides is that specific compounds can be prepared to test given hypotheses. The number of competing effects can be kept reasonably small, and the number of possible conformations is a much more tractable quantity. In Chapters III and IV, we shall show that the bases in dinucleoside phosphates and trinucleoside diphosphates have similar conformations. These compounds have optical properties which are very similar to single strand polynucleotides. Thus they are suitable models for RNA, and should be submitted to further study. In contrast, the mononucleosides or mononucleotides are very poor models for polymers. It remains a future task to obtain and study suitable oligonucleotide models for double stranded polynucleotides, though some preliminary experiments and ideas will be discussed in Chapter IV.

One of the most intriguing properties of RNA is the existence of many sequence dependent phenomena. Oligonucleotides can serve as excellent models for studying these effects. Best studied thus far has been the degradation of oligomers by sequence specific endonucleases like pancreatic ribonuclease.[261] We fully expect that future studies will produce sequence specific chemical reagents. Some evidence has accumulated that the absorption spectrum of oligonucleotides is sequence dependent.[151] Recent work in this laboratory has amplified the number of examples,[252] and additional evidence will be discussed in Chapter III. In addition, we shall show that the sequence dependence of the ORD is

much more dramatic than absorbance changes. Our optical studies have been greatly helped by the very intense absorption bands found in nucleic acid bases. These permit studies which employ very small amounts of material. Lack of suitable quantities of oligonucleotides has thus far impeded extensive titrimetric or NMR studies of these compounds. But with the advent of more efficient methods of preparation, such studies will almost certainly be forthcoming.[223,121] All of the effects discussed so far are due directly to the sequence of bases on a single strand. Another class of sequence dependent properties should arise from the strong sequence dependence of molecular conformation. All of the effects discussed above should be sensitive to conformation. In addition, viscosity, sedimentation, and other hydrodynamic phenomena should be dependent on sequence. The anomalous temperature dependence of the viscosity of poly d(AT) when compared with most DNA's is a well characterized example of the extreme effects possible with a polynucleotide of unusual sequence.[103,215] It remains to be seen whether the conformation of RNA's is so delicately balanced that a small change in sequence may effect the overall properties of a large macromolecule. But it is almost certain that sequence dependent phenomena play an important role in determining the biological properties of RNA.

II. EXONUCLEASE KINETICS

1. Methods of Determining Nucleotide Sequence

The motivations for determining the base sequences of nucleic acids are very strong. In the case of information carrying macromolecules like DNA and mRNA, a knowledge of the nucleotide sequence will enable us to decipher the genetic message encoded on these nucleic acids. Establishing the base sequence of rRNA and sRNA is an essential step in the determination of the conformation of these molecules, and will be of great help in understanding their function. Thus there has been much work in the past few years on the problem of determining nucleic acid sequences.

Procedures for the determination of the amino acid sequence in proteins have been available for some time, and the primary structures of many have yielded to these approaches. However, the problem of ascertaining the base sequence of all but the smallest nucleic acids remains formidable. Two recent reviews have summarized many of the possible methods for the determination of nucleic acid sequence.[24,51] It is quite apparent that the general problem of obtaining a nucleic acid sequence is much more difficult than for a protein sequence, because most nucleic acids are so much larger. Also, there are many more different types of residues in a given protein than in an average nucleic acid chain of the same length. While some of the basic techniques which have proven successful on proteins have been adapted for nucleic acids, it is certainly necessary that new and more powerful methods be found. In this chapter we review some of the methods that are currently in use and discuss some of the proposed improvements. Then we will explore in

detail the possibilities for using exoenzymes for obtaining sequence information. This is a method which may become an important adjunct to many of the others that have been proposed.

By far the most popular method for determining nucleic acid sequences is the overlap method first used by Sanger on insulin. Using this method Holley and his coworkers have been successful in determining the base sequence of the yeast alanine sRNA, a polymer, which contains 77 nucleotides.[97] Many other laboratories are proceeding with similar studies on other sRNA's,[4,11,43,267] though none of these attempts have yet yielded a complete sequence. Holley first hydrolyzed the RNA with two specific endonucleases, T1 ribonuclease (EC 3.1.4.8) and pancreatic ribonuclease (EC 2.7.7.16). The mixture of oligonucleotides from each digestion was chromatographically separated, and the products identified.[98] A hypothetical example is:

$$\text{ApGpUpCpGp} \xrightarrow{\text{Panc. RNAase}} \text{ApGpUp} + \text{Cp} + \text{Gp}$$

$$\text{ApGpUpCpGp} \xrightarrow{\text{T1 RNAase}} \text{ApGp} + \text{UpCpGp}$$

This permits the sequence of the pentanucleotide to be reconstructed. Then partial hydrolyses were used to obtain larger fragments which were in turn further degraded and identified.[2,168] These provided the overlaps necessary to reconstruct the entire sequence. Of great help was the fact that this RNA contained 11 different bases instead of the more usual four. It will be difficult to use this method on much larger polymers. The major problems will be separating the fragments formed and finding conditions under which more carefully controlled partial hydrolyses can be carried out. There have been many imaginative attempts

made to improve our ability to separate oligonucleotides.[73,221,44] But there does not yet exist a separation technique capable of resolving the long sequences needed to determine the sequence of a large viral RNA. Nevertheless, Holley's method is very satisfactory for handling nucleic acid chains of the order of 100 bases long, and presently the most practical technique of any that have been suggested.

A variation of the simple overlap method was proposed by Rice and Bock.[180] They suggested that the troublesome partial hydrolyses can be eliminated if a sufficient number of different specific endonucleases can be found. In most cases the analysis of the fragments produced by four total degradations of RNA, each with an enzyme specific for cleaving after a different one of the four bases A, C, U, and G, would lead to the sequence. There are currently available two enzymes with such specificities. T1 RNAase cleaves specifically after G,[51] and E. coli Ribosomal RNAase is apparently quite specific for C.[1] It is also possible to modify RNA in such a way that pancreatic RNAase becomes specific for C.[71] There is some evidence that T2 RNAase shows some preference to cleave after A.[51] Thus while our repertoire of nucleases is not yet large enough to permit an experimental test of the method proposed by Rice and Bock, enzymes with the necessary specificities will probably soon be found.

Another improvement on the overlap method has been discussed by Mandeles and Tinoco.[144] This involves labeling one end of the nucleic acid with some distinguishable marker. The RNA is then partially hydrolyzed enzymatically and all the labeled fragments are separated according to chain length, isolated and purified. These are then labeled with a different marker on the other end, and hydrolyzed to completion

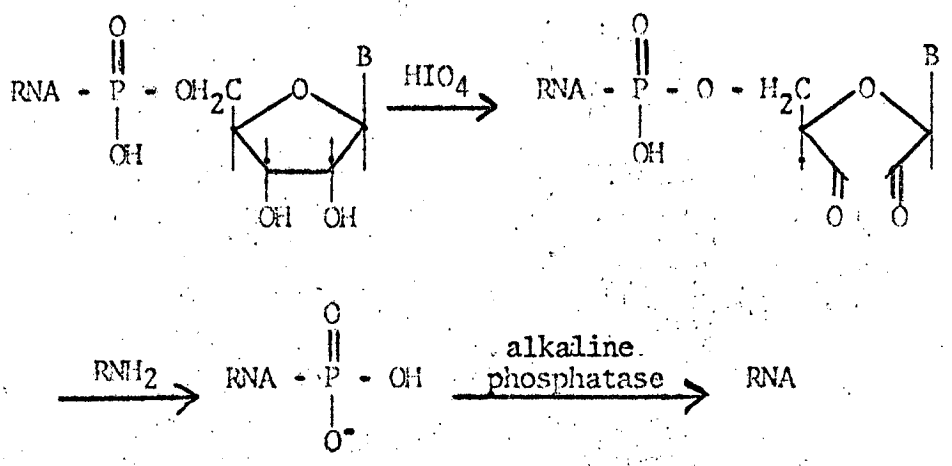
by the same specific enzyme previously used. The fragments labeled with the second marker are then isolated and identified. This permits ordering all of the fragments, and thus eliminates the need for the determination of overlaps. The difficulties encountered here are the lack of a suitable label, and the difficulty of chromatographically separating very long chain lengths. This method is presently being used by Mandeles to determine the 3' terminal sequences of TMV RNA.[143] Similar end labels have proven very useful in the discovering of the 3' terminal sequence of several sRNA's.[10,114]

An entirely different approach is the use of electron microscopy. While the individual bases are too similar to distinguish with the present resolution, Moudrianakis and Beer[9] have found that it is possible to find chemical reagents which attach to DNA and can then be clearly seen along the DNA strand in the electron microscope. Thus far they have found a G specific marker,[156,157] and some evidence for a marker specific for C.[149] If markers can be found that are equally specific for T or U and A, then this method should provide the simplest and most effective way of determining the sequence of large nucleic acids.

Another method which offers great promise but which defeats part of the purpose of determining nucleic acid sequence is to work back from known protein sequences using the genetic code. Undertainties caused by the degeneracy of codons could easily be solved by using mutations which change the phase of the DNA, such as deleting or adding one nucleotide,[36,215] or simple mutations from one amino acid to another.[112,243] Genetic mapping on a slightly larger scale could be extremely valuable in ordering sections of a DNA whose partial sequence has been determined by more classical methods. The limitations of genetic mapping are discussed in the

recent book by Haggis, et al.[82] These techniques merit consideration because it is fundamentally easier to determine the sequence of proteins than nucleic acids.

The chemical degradation of a nucleic acid step by step is conceptually a very simple way of determining the sequence. An effective way of doing this, described by Yu and Zamecnik,[265] is shown below.



Recent improvements on the same general scheme are discussed by Neu and Heppel.[159] The major difficulty is that there are at least six steps involved, not all of which proceed with 100% yield. Thus the results after many degradations will be random. This method is most applicable for determining the 3' terminal nucleotide, although other simple and effective methods for this exist.[122] If terminal degradation can be automated, and if the yields of each step can be improved, it will become a fine method of sequence determination. Considerable progress has been made in this direction.[138,225]

A logical extension to the method of terminal chemical degradation is to obtain sequence information using exoenzymes. These enzymes catalyze the reaction shown schematically below.

$$ABCDE \longrightarrow ABCD + E \longrightarrow ABC + D + E, \text{ etc.}$$

The advantage of exonucleases over terminal chemical reactions is that the former work at much higher yield. The disadvantage is that there is no controlled way of taking off only one monomer at a time. Several preliminary experimental results have been obtained by Nihei and Cantoni,[161] and by Singer and Fraenkel-Conrat,[205,64] but these are difficult to interpret due to such problems as lack of homogeneity of the starting polymer[206] and impurity of the enzymes. In the following sections we will discuss in detail the problem of obtaining sequence information by studying the products of an exonuclease digestion of RNA. We will calculate the kinetics of action of exonucleases using several simplified models. This will provide a framework for the application of kinetic data on suitable model substrates to the determination of the monomeric sequences of polymers.

2. Properties of Exonucleases

A number of exonucleases have been obtained in recent years in various degrees of purity. Some basic properties of these enzymes are summarized in Table I. Three general types of exonucleases have been characterized. First discovered were 5' phosphodiesterases, which degrade an RNA or DNA by the following scheme.

$$ApBpCpD \longrightarrow ApBpC + pD \longrightarrow ApB + pD + pC \longrightarrow A + pB + pC + pD$$

Thus all of these enzymes start from the free 3' end, and yield 5' nucleotides as the major product. The earliest enzyme discovered and the most fully studied is venom phosphodiesterase (EC 3.1.4.1) (VPDE).[111,173,258,177,13,109] VPDE can degrade both RNA and DNA, but RNA oligomers are hydrolyzed several times slower than DNA oligomers.[177] The presence of a

TABLE I

Summary of the Properties of Some Exonucleases

Enzyme	Substrates	End Susceptible	Major Product	Base Specificity	Conformation Specificity
Venom PDE[177]	RNA & DNA	3'*	5' Nucleotide	Slight	Questionable
Spleen PDE[179]	RNA & DNA	5'*	3' Nucleotide	Unknown	Some
L. acidophilus[59]	RNA & DNA	5'*	3' Nucleotide	None	None
B. subtilis[110]	RNA & DNA	5'	3' Nucleotide	Unknown	Unknown
L. casei[108]	RNA	3'	5' Nucleotide	Unknown	Unknown
E. coli RNAase II[210]	RNA	3'	5' Nucleotide	Needs A,U, C or G	Single strands
E. coli DNAase I[125]	DNA	3'*	5' Nucleotide	Unknown	Denatured DNA
E. coli DNAase II[125]	DNA	3'*	5' Nucleotide	Unknown	Prefers native DNA
E. coli DNAase III[125]	DNA	3'	5' Nucleotide	Unknown	Native DNA
M. lysodeikticus[79]	RNA	3'	5' Nucleoside-diphosphate	None	Single strands

*known to require free terminal OH

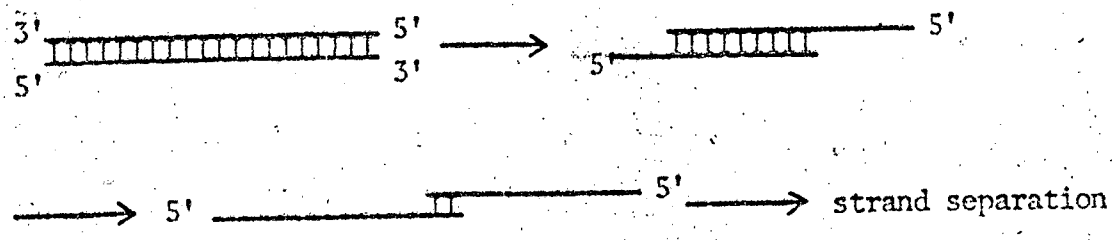
5' phosphate accelerates the rate of digestion of an oligomer by up to a factor of 20, although this unusual effect presumably will decrease as a function of chain length.[177] A 3' terminal phosphate drastically slows the rate of enzyme degradation,[173] as does a 3'-O-acetyl.[178] Most evidence seems to indicate that there is little base specificity or conformational specificity,[161,162] although this is by no means certain.[109] This enzyme can easily be obtained relatively free of endonuclease, but most preparations are usually contaminated with a 5' nucleotidase.[109,13] VPDE is the exoenzyme that has been used most often in attempts to determine the sequence of RNA's.

Unlike VPDE, which readily degrades oligonucleotides, RNAase II from E. coli (EC 3.1.4.1) is specific for chain lengths greater than five or six.[209,210] It also shows strong preference for single stranded RNA, and cannot degrade DNA. RNAase II shows little preference among the four common nucleotide bases, but it cannot degrade 4-N-methyl C or 4-N, N-dimethyl C. Poly I is also completely resistant, although this may be due to the presence of a 3 strand helix under the conditions where the enzyme was tried.

The 5' phosphodiesterase from L. casei[108] (EC 3.1.4.1) appears to be quite similar to RNAase II. One unusual property of this enzyme is that it degrades RNA to chains of intermediate length (12-24) and then degrades this oligomer all at one to monomers. Also, this enzyme cannot degrade Worthington yeast RNA, but has no trouble with most other common RNA's. There is some suspicion that the 5' phosphodiesterase from E. coli and L. casei may be responsible for the breakdown of mRNA in vivo.[210]

Three interesting 5' phosphodiesterases have been found in E. coli by Lehman.[125,126] These are all specific for DNA, although DNAase I and

II will break an RNA link incorporated into DNA. The presence of a free 3' end is essential for both DNAase I and II to act. DNAase I is specific for denatured DNA by a factor of more than 40,000. It cannot degrade a dimer to monomers, and attacks oligomers of length 3 to 6 faster than polymers. In contrast, DNAase II (EC 3.1.4.6) can hydrolyze dimers, but it prefers native DNA, which it hydrolyzes several times faster than denatured DNA. DNAase III is specific for native DNA. DNA is hydrolyzed only to about 40% of completion, which is consistent with the proposed mechanism shown below.



DNAase III can hydrolyze poly d(AT) essentially to completion, which is consistant with what is known about the structure of that polymer. It has been difficult thus far to characterize fully the activity of this enzyme, due to the presence of phosphatase activity in even the most purified DNAase III fractions.

The second kind of exonucleases are the 3' phosphodiesterases. They degrade RNA from the 5' end to yield 3' nucleotides.

$$ApBpCpD \longrightarrow Ap + BpCpD \longrightarrow Ap + Bp + CpD \longrightarrow Ap + Bp + Cp + D$$

The first 3' phosphodiesterase studied was obtained from spleen.[94,179] This enzyme shows no base specificity, and hydrolyzes RNA oligomers and DNA oligomers at similar rates. Spleen phosphodiesterase (EC 3.1.4.1) needs a free 5' end. Most preparations have very low endoenzyme im-

purity,[32] and show some conformational specificity. A complicating factor in the use of spleen phosphodiesterase is that the enzyme can catalyze synthetic reactions under the conditions commonly used.[179] Care must be taken not to let the concentrations of product build up.

Another 3' phosphodiesterase has been obtained from B. subtilis[110] (EC 3.1.4.1). It has not yet been studied in detail, but it is known that this enzyme attacks RNA at about half the rate of DNA, and that the enzyme requires Ca^{++}. The third enzyme in this category was obtained from L. acidophilus[59] (EC 3.1.4.1). This enzyme degrades both TpT and UpA at the same rate. It also seems to show no kinetic specificity in degrading p-nitrophenyl deoxynucleotides. The presence of a 5' terminal phosphate slows the rate of attack by more than a factor of 50, and 3' methyl nucleotides are completely resistant to attack. The rate of hydrolysis of oligomers decreases strongly with chain length at pH 7, but not at pH 4.5. Thus, polymers at pH 7 are not degraded at any appreciable rate.[60] There is no apparent conformational specificity, but this would be hard to measure at pH 7 due to the effect of chain length. The endo activity in preparations of this enzyme is less than 7/10,000 of the exo activity. Thus this is a very desirable enzyme for use in sequence studies, where, as will be shown later, the effect of an endoenzyme impurity is very damaging.

The third class of exoenzymes are the polynucleotide phosphorylases.[79] The most well studied of these is obtained from M. lysodeikticus[207] (EC 2.7.7.8). This enzyme is only an effective exonuclease in the presence of a large excess of inorganic phosphate. It acts according to the following scheme, yielding 5' diphosphates as products.

$$ApBpCpD \xrightarrow{P_i} ApBpC + ppD \xrightarrow{P_i} ApD + ppC + ppD \xrightarrow{P_i} A + ppB + ppC + ppD$$

The M. lysodeikticus enzyme degrades single strands much more rapidly than double strands, and has no base specificity. In the limit of high enzyme concentration there is no effect of chain length. This enzyme also has the advantage of very low endonuclease contaminants. See Appendix 6 for further discussion of the preparation and properties of this enzyme.

A major question about exonucleases remains unanswered. Thus far all enzymes which degrade from the 5' end yield 3' phosphorylated products, and those which start from the 3' end produce 5' phosphorylated products. Whether this is a fact of nature or merely the result of an incomplete survey of enzymes remains to be seen.

To simplify the rest of the discussion in this chapter, we shall make several assumptions about the nature of exonucleases and their substrates. These assumptions are for the most part generalizations of the properties we have just discussed.

1. We shall deal with enzymes that have no base specificity. Thus, the overall rate of liberation of monomers is independent of sequence, or base composition.

2. We shall assume that the kinetics are independent of chain length. This implies that the terminal phosphates of the nucleic acid have been removed, and that the initial rate of degradation of oligomers is independent of chain length.

3. Our substrates are single stranded polymers. Or, alternatively, our nucleases have no dependence on the polymer conformation.

Suppose we had only one polymer molecule and one exoenzyme in solution which worked according to the above assumptions. If we monitored the base

composition of the solution at sufficiently small time intervals we would identify each succeeding monomer as it was liberated, and thus we could determine the entire sequence of the polymer. In practice, such an experiment lies 10 to 15 orders of magnitude below our current limits of detection. Thus, we must be content in using a population of polymers and enzymes. We could predict the base composition of the ensuing mixture as a function of time, either by statistics, or by solving the differential equations for the kinetics of such a reaction. It is the latter approach which we have chosen to use.

3. Kinetics of Random Attack

The first kinetic scheme which we will consider corresponds to the irreversible degradation of a chain from one end, in a sequence of steps as indicated below.[27]

$$\begin{aligned}
&E + S_N \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES_N \xrightarrow{k_2} E + S_{N-1} + B_1 \\
&\vdots \\
&E + S_{N-n+1} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES_{N-n+1} \xrightarrow{k_2} E + S_{N-n} + B_n \qquad (1) \\
&\vdots \\
&E + S_2 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES_2 \xrightarrow{k_2} E + S_1 + B_{N-1}
\end{aligned}$$

Since this scheme allows enzyme molecules to attack substrate molecules at random, we shall call this the random model. S_N refers to the undegraded nucleic acid chain, and B_1 to the first nucleotide liberated. Only a single enzyme-substrate complex, ES_{N-n}, is involved for each degradation step. Undoubtedly this is an oversimplification for these

scission reactions, but at the present stage of knowledge the added refinement of multiple enzyme-substrate complexes seems unwarranted.

The kinetics of the above scheme can be represented by a set of differential equations for (a) the polymeric chains,

$$(\dot{S}_N) = \frac{d(S_N)}{dt} = -k_1(E)(S_N) + k_{-1}(ES_N)$$

$$(\dot{S}_{N-n}) = -k_1(E)(S_{N-n}) + k_{-1}(ES_{N-n}) + k_2(ES_{N-n+1}) \qquad (2a)$$

$$(\dot{S}_1) = k_2(ES_2)$$

(b) the enzyme-substrate complexes,

$$(\dot{ES}_{N-n}) = k_1(E)(S_{N-n}) - (k_{-1} + k_2)(ES_{N-n}) \qquad (2b)$$

and (c) the literated nucleotides.

$$(\dot{B}_{N-1}) = k_2(ES_{N-n+1}) \qquad (2c)$$

The final step produces two nucleotides, B_{N-1} and $S_1 = B_N$. In this approximation the last two monomers are kinetically indistinguishable. A more realistic model for the kinetics of exonuclease degradation would have a final step where B_{N-1} is released either before or after B_N.

$$E + S_2 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES_2 \xrightarrow{k_2} ES_1' + B_2 \xrightarrow{k_2'} E + B_1 \qquad (3)$$

We have not used this model since it would be hard to justify using multiple enzyme-substrate complexes for only the last step. Instead, we can make the approximation that B_N is liberated from a hypothetical S_1, or, instead, avoid consideration of the last two bases of the polymer.

None of the approximations made for the last steps of the reaction can affect the kinetics of the earlier steps in this model.

An analytical solution of the above set of non-linear differential equations is not attainable. However, with the approximation that the concentration of free enzyme, E, is substantially independent of time, the equations become linear and a solution can be effected. Such a condition can correspond physically to a situation where the total enzyme concentration greatly exceeds the concentration of reactive ends. As it is only in the last stages of the degradation that the number of reactive chain ends diminishes, the approximation of a constant level of free enzyme should apply throughout nearly all of the reaction. It will be shown later that the assumption of constant enzyme concentration is quite good for a typical choice of rate constants and initial substrate and enzyme concentrations.

The differential equations, 2a, b, and c, are linearized by assuming that E is a constant. $k_1(E) = k$. The boundary conditions are

$$S_{N-n}(0) = B_n(0) = ES_{N-n}(0) = 0 \qquad n \neq 0$$

$$B_N(0) = 0 \; ; \; S_N(0) \text{ and } ES_N(0) \neq 0$$

The Laplace transform of a function, F(t) is

$$\mathcal{L}(F(t)) = \int_0^\infty e^{-st} F(t)\, dt$$

Define the following Laplace transforms:

$$\begin{aligned} \mathcal{L}(ES_{N-n}) &= \overline{ES}_{N-n} \\ \mathcal{L}(S_{N-n}) &= \overline{S}_{N-n} \\ \mathcal{L}(B_n) &= \overline{B}_n \end{aligned} \qquad (4)$$

By taking the Laplace transform of equations 2a, b, and c, and using the result that $(F'(t)) = s\mathcal{L}(F(t)) - F(0)$ we have,

$$s\bar{S}_N - S_N(0) = -k\bar{S}_N + k_{-1}\overline{ES}_N \tag{5a}$$

$$s\bar{S}_{N-n} = -k\bar{S}_{N-n} + k_{-1}\overline{ES}_{N-n} + k_2\overline{ES}_{N-n+1} \qquad n \neq 0, N \tag{5b}$$

$$s\overline{ES}_N - ES_N(0) = k\bar{S}_N - (k_{-1} + k_2)\overline{ES}_N \tag{5c}$$

$$s\overline{ES}_{N-n} = k\bar{S}_{N-n} - (k_{-1} + k_2)\overline{ES}_{N-n} \tag{5d}$$

$$s\bar{B}_n = k_2\overline{ES}_{N-n+1} \tag{5e}$$

This set of linear equations is solved algebraically. We find that

$$\bar{B}_n = \frac{k_2(kk_2)^{n-1}ES_N(0)}{[s^2+s(k+k_{-1}+k_2)+kk_2]^n} + \frac{(kk_2)^n[ES_N(0)+S_N(0)]}{s[s^2+s(k+k_{-1}+k_2)+kk_2]^n} \tag{6}$$

The denominator can be factored into terms of the form $(S - \alpha_1)^n$ and $(s - \alpha_2)^n$, where α_1 and α_2 are roots of a quadratic equation.

$$\alpha_{1,2} = \frac{-(k + k_{-1} + k_2) \pm \sqrt{(k + k_{-1} + k_2)^2 - 4\,kk_2}}{2} \tag{7}$$

Using the fact that $kk_2 = \alpha_1\alpha_2$, we have

$$\bar{B}_n = \frac{k_2(\alpha_1\alpha_2)^{n-1}ES_N(0)}{(s-\alpha_1)^n(s-\alpha_2)^n} + \frac{(\alpha_1\alpha_2)^n[ES_N(0)+S_N(0)]}{s(s-\alpha_1)^n(s-\alpha_2)^n} \tag{8}$$

Taking the inverse transform of equation 8, we find

$$B_n = k_2(\alpha_1\alpha_2)^{n-1}ES_N(0)\,\mathcal{L}^{-1}\left(\frac{1}{(s-\alpha_1)^n(s-\alpha_2)^n}\right) + (\alpha_1\alpha_2)^n[ES_N(0)+S_N(0)]\,\mathcal{L}^{-1}\left(\frac{1}{s(s-\alpha_1)^n(s-\alpha_2)^n}\right) \tag{9}$$

It is known that

$$\mathcal{L}^{-1}[(s-\alpha_1)^{-1}(s-\alpha_2)^{-1}] = (\alpha_1-\alpha_2)^{-1}(e^{\alpha_1 t}-e^{\alpha_2 t}). \quad 29$$

By successive applications of $\partial^2/\partial\alpha_1\partial\alpha_2$ to both sides of this equation, we find that

$$I_n \equiv \mathcal{L}^{-1}\left(\frac{1}{(s-\alpha_1)^n(s-\alpha_2)^n}\right)$$

$$= \frac{1}{(n-1)!}\sum_{i=0}^{n-1}\frac{t^i(2n-2-i)!(-1)^{n-1+i}[e^{\alpha_1 t}+(-1)^{i+1}e^{\alpha_2 t}]}{(n-1-i)!(i)!(\alpha_1-\alpha_2)^{2m-1-i}} \qquad (10)$$

By similar methods, after performing the change of variables $v_1 = \alpha_1$, $v_2 = \alpha_1-\alpha_2$, and using the expansion

$$\left(\frac{\partial^2}{\partial\alpha_1\partial\alpha_2}\right)^m = (-1)^m\sum_{i=0}^{m}\left(\frac{\partial^2}{\partial v_2^2}\right)^{m-i}\left(\frac{\partial^2}{\partial v_1\partial v_2}\right)^i\frac{m!}{(m-i)!(i)!}$$

we can show that

$$II_n \equiv \mathcal{L}^{-1}\left(\frac{1}{s(s-\alpha_1)^n(s-\alpha_2)^n}\right) = \frac{1}{(\alpha_1\alpha_2)^n} +$$

$$\left(\frac{(-1)^{n-1}}{(n-1)!}\sum_{j=1}^{2}\sum_{i=0}^{n-1}\frac{(2n-2-i)!}{(n-1-i)![(-1)^{j+1}\alpha_1+(-1)^j\alpha_2]^{2n-i-i}}\right)X \qquad (11)$$

$$\left(\sum_{p=0}^{i}\frac{(-1)^p t^p e^{\alpha_j t}}{(p)!(\alpha_j)^{i+1-p}}\right)$$

Using the above set of sums, we rewrite equation 9 as

$$B_n(t) = \frac{ES_N(0)}{k} (\alpha_1\alpha_2)^n \prod_n + [ES_N(0)+S_N(0)](\alpha_1\alpha_2)^n \coprod_n \quad (12)$$

The expressions for the concentration of the liberated nucleotides are in general polynomial sums in powers of t multiplied by two exponents, $e^{\alpha_1 t}$ and $e^{\alpha_2 t}$. The presence of only two distinct exponential factors is a consequence of the assumption of a single enzyme-substrate complex at each step. Additional complexes would introduce more rate constants and exponential factors.

4. Steady state and optimal approximations

Attention will be focused on two particular experimental situations. The first condition corresponds to a steady state concentration of the various enzyme-substrate complexes. Such a condition applied if $k \ll k_{-1}$ or k_2, or if the reverse inequality holds. Then expansion of the square root in equation (7) with retention of the first two terms yields

$$\alpha_1 \cong - \frac{kk_2}{k + k_{-1} + k_2} \quad (13a)$$

and

$$\alpha_2 \cong - (k + k_{-1} + k_2) . \quad (13b)$$

Consider the well known Michaelis-Menten model for the kinetics of enzyme action.[69] For the reaction

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} P + E$$

the dissociation constant of ES is defined as K_m, the Michaelis constant,

$$K_m = \frac{k_{-1} + k_2}{k_1} = \frac{[(E(0) - ES)][S]}{[ES]} \quad (14)$$

where E(O) is the total enzyme concentration. The maximum velocity of the reaction occurs when all of the enzyme is bound to the substrate

$$V_m = k_2 \quad [E(O)] \quad . \tag{15}$$

Thus, using the assumption of a steady state level of complex, the familiar result is obtained that

$$\frac{d[P]}{dt} = k_2 \, [ES] = \frac{V_m \, [S]}{K_m + [S]} \quad .$$

Likewise, it is to be noted that the parameter α_1 in our exonuclease kinetics can be written in terms of the customary parameters of steady state enzyme kinetics

$$\alpha_1 = \frac{-kk_2}{k + k_{-1} + k_2} = \frac{-k_2}{1 + \frac{k_{-1} + k_2}{k_1 E}} = \frac{-V_m / [E(O)]}{1 + K_m / [E]} \tag{16}$$

The steady state results are in fact independent of mechanistic details, as the inclusion of additional enzyme substrate complexes at each step would only alter the definitions of V_m and K_m.[167] As $|\alpha_2| >> |\alpha_1|$, the steady state terms involving $e^{\alpha_1 t}$ die out much more rapidly than the steady state terms involving $e^{\alpha_1 t}$.[154,5]

A solution corresponding to the steady state condition for all complexes, $(ES_{N-n}) = 0$, is also presented for the purpose of comparative calculations.

In the steady state approximation equations (5c) and (5d) are equal to zero. By algebraic manipulation as before, we have that

$$\overline{B}_n = \frac{\beta^n S_N(O)}{s(s+\beta)^n} \quad \text{where} \quad \beta = \frac{kk_2}{k_{-1} + k_2} \tag{17a}$$

Taking the inverse transform of (17a) we find

$$B_n = \beta^n S_N(0) \mathcal{L}^{-1}\left(\frac{1}{s(s+\beta)^n}\right) . \tag{17b}$$

By methods as before we can show that

$$\mathcal{L}^{-1}\left(\frac{1}{s(s+\beta)^n}\right) = \frac{1}{\beta^n}\left(1 + \sum_{i=0}^{n-1} \frac{(\beta t)^i e^{-\beta t}}{i!}\right) , \tag{18}$$

Hence, the steady state solution for the concentration of a given nucleotide is found to be given by

$$B_n = S_N(0)\left[1 - \sum_{i=0}^{n-1} \frac{(\beta t)^i e^{-\beta t}}{i!}\right] \tag{19}$$

The second case of interest occurs when (a) k and $k_2 \gg k_{-1}$, and (b) $k \equiv k_1 (E) \cong k_2$. Then from equation (7)

$$\alpha_1 \cong \alpha_2 \cong -k_2 \equiv -\alpha$$

In this special case, equation (11) does not hold. Instead,

$$\prod_n = \mathcal{L}^{-1} \frac{1}{s(s+\alpha)^{2n}}$$

and using equation (9) and the assumption that $ES_N(0) = 0$, we have that the concentration of the nucleotides is given by

$$B_n(t) = S_N(0)\left[1 - \sum_{i=0}^{2n-1} \frac{(\alpha t)^i e^{-\alpha t}}{i!}\right] . \tag{20}$$

Conditions (a) and (b) minimally require a total enzyme concentration of at least the same order of magnitude as K_m inasmuch as

$$\frac{K_m}{E(0)} = \frac{k_{-1} + k_2}{k_1 E(0)} \leq \frac{k_2}{k}$$

Plainly in this situation there is no distinction between the transient and steady state phases of each reaction step. The steady state case and the optimal case ($\alpha_1 = \alpha_2$) represent the extremes of behavior of the random model of attack.

By using the original differential equations, expressions for $S_{N-n}(t)$ and $ES_{N-n}(t)$ can be derived. In the steady state case these are

$$ES_{N-n+1}(t) = \frac{\beta S_N(0)(\beta t)^{n-1} e^{-\beta t}}{k_2(n-1)!}, \quad S_{N-n+1}(t) = \frac{S_N(0)(\beta t)^{n-1} e^{-\beta t}}{(n-1)!} \tag{21}$$

and in the case $\alpha_1 = \alpha_2 = -\alpha$ we have

$$ES_{N-n+1}(t) = \frac{S_N(0)(\alpha t)^{2n-1} e^{-\alpha t}}{(2n-1)!}, \quad S_{N-n+1}(t) = \frac{S_N(0)(\alpha t)^{2n-2} e^{-\alpha t}}{(2n-2)!} \tag{22}$$

The values of S_{N-n+1} for the steady state case are in complete agreement with the results calculated by Flory for the analogous condensation of ethylene oxide polymers.[63]

Another quantity of interest is the total rate at which monomer is evolved from an infinite polymer. This is the slope of a plot of total monomer concentration versus time. In the steady state

$$\sum_{N=1}^{\infty} \dot{B}_n(t) = S_N(0)\beta \sum_{n=1}^{\infty} \frac{(\beta t)^{n-1} e^{-\beta t}}{(n-1)!} = \beta S_N(0) \quad . \tag{23}$$

and in the case $\alpha_1 = \alpha_2 = -\alpha$

$$\sum_{n=1}^{\infty} \dot{B}_n(t) = S_N(0)\alpha \sum_{n=1}^{\infty} \frac{(\alpha t)^{2n-1} e^{-\alpha t}}{(2n-1)!} = \alpha(1/2 - 1/2e^{-2\alpha t}) S_N(0) \tag{24}$$

Thus we can see that my measuring the total rate at which monomer is liberated it may be possible to determine the kinetic parameters α and β, and perhaps to decide whether or not an exoenzyme is acting under steady state conditions.

In most conventional kinetic studies, $K_m \gg E(0)$, so that

$$\alpha_1 \cong \frac{V_m E}{K_m E(0)} = \frac{-k_2 k_1 E(0)}{k_{-1} + k_2} \left[\frac{E}{E(0)} \right]$$

But according to our assumptions $E/E(0) \cong$ constant $\cong 1$. Therefore $\alpha_1 \cong -\beta$ under these conditions, and it can be seen that the complete model will approach the steady state case in the appropriate limit.

From recent studies on a phosphodiesterase from L. acidophilus,[59] K_m is found to be rather high (ca. 10^{-3} M) for various synthetic substrates so that the above condition would be expected to apply.

To examine these points more quantitatively, calculations (employing an IBM 7090 digital computer) have been performed with various values of the kinetic parameters to determine the time course of appearance of successive nucleotides. The details of the computer programs used are given in Appendix 1.

The values of $B_n(t)$ for the steady state case are collected in Table II for n = 1 to 14. These values can also be used in the case $\alpha_1 = \alpha_2$, since there $B_n(\alpha t)$ equals the steady state $B_{2n-1}(\beta t)$. As will be seen later, the values of $B_n(t)$ in Table II will also be of use in describing the kinetics of another model of exoenzyme activity.

TABLE II

Values of $B_n(\beta t)/S_N(0)$ for the Steady State Case of the Random Model

Time	n													
βt	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
0.5	.394	.090	.014	.002	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
1.0	.632	.264	.080	.019	.004	.001	.000	.000	.000	.000	.000	.000	.000	.000
1.5	.777	.442	.191	.066	.019	.005	.001	.000	.000	.000	.000	.000	.000	.000
2.0	.865	.594	.323	.143	.053	.017	.005	.001	.000	.000	.000	.000	.000	.000
2.5	.918	.713	.456	.242	.109	.042	.014	.004	.001	.000	.000	.000	.000	.000
3.0	.950	.801	.577	.353	.185	.084	.034	.012	.004	.001	.000	.000	.000	.000
3.5	.970	.864	.679	.463	.275	.142	.065	.027	.010	.003	.001	.000	.000	.000
4.0	.982	.908	.762	.567	.371	.215	.111	.051	.021	.008	.003	.001	.000	.000
4.5	.989	.939	.826	.658	.468	.297	.169	.087	.040	.017	.007	.002	.001	.000
5.0	.993	.960	.875	.735	.560	.384	.238	.133	.068	.032	.014	.006	.002	.001
5.5		.973	.912	.798	.643	.471	.314	.191	.106	.054	.025	.011	.005	.002
6.0		.983	.938	.849	.715	.554	.394	.256	.153	.084	.043	.020	.009	.004
6.5		.989	.957	.888	.776	.631	.474	.327	.208	.123	.067	.034	.916	.007
7.0		.993	.970	.918	.827	.699	.550	.401	.271	.170	.099	.053	.027	.013
7.5			.980	.941	.868	.759	.622	.475	.338	.224	.138	.079	.043	.022
8.0			.986	.958	.900	.809	.687	.547	.408	.283	.184	.112	.064	.034

TABLE II (continued)

Time	n													
βt	1	2	3	4	5	6	7	8	9	10	11	12	13	14
8.5			.991	.970	.926	.850	.744	.614	.477	.347	.237	.151	.091	.051
9.0				.979	.945	.884	.793	.676	.544	.413	.294	.197	.124	.074
9.5				.985	.960	.912	.835	.731	.608	.478	.355	.248	.164	.102
10.0				.990	.971	.933	.870	.780	.667	.542	.417	.303	.208	.136
10.5				.993	.979	.950	.898	.822	.721	.603	.479	.361	.258	.175
11.0					.985	.963	.921	.857	.768	.660	.540	.421	.311	.219
11.5					.989	.972	.940	.886	.809	.711	.598	.480	.367	.267
12.0					.992	.980	.954	.911	.845	.758	.653	.538	.424	.319
12.5						.985	.965	.930	.875	.799	.703	.594	.481	.372
13.0						.989	.974	.946	.900	.834	.748	.647	.537	.427
13.5						.992	.981	.959	.921	.865	.789	.696	.591	.482
14.0							.986	.968	.938	.891	.824	.740	.642	.536
14.5							.990	.976	.952	.912	.855	.780	.689	.588
15.0							.992	.982	.962	.930	.881	.815	.732	.637
15.5								.987	.971	.945	.904	.846	.772	.683
16.0								.990	.978	.957	.923	.873	.807	.726
16.5									.983	.966	.938	.896	.838	.764
17.0									.987	.974	.951	.915	.865	.799

5. Validity of Constant Enzyme Approximation

Before discussing the results of the calcuations, we would like to show that the assumption of constant enzyme concentration is a reasonable one. By conservation of mass we have that

$$[E] = E(0) - \sum_{n=1}^{N} E\, S_{N-n+1} \tag{25a}$$

and

$$\frac{d[E]}{dt} = - \sum_{n=1}^{N} \frac{d[E\, S_{N-n+1}]}{dt} \tag{25b}$$

In the early stages of the reaction we can write a similar equation for the total concentration of substrate.

$$\sum_{n=1}^{N} [S_{N-n+1}] = S_N(0) - \sum_{n=1}^{N} [ES_{N-n+1}] \tag{26}$$

From the proposed reaction scheme we have already shown that

$$\frac{d[ES_{N-n+1}]}{dt} = k_1[E][S_{N-n+1}] - (k_{-1} + k_2)\, [E\, S_{N-n+1}] \cdot$$

Substituting equation (25b) into the above we have that

$$\frac{d[E]}{dt} = \sum_{n=1}^{N} \left\{ k_1[E][S_{N-n+1}] - (k_{-1} + k_2)\, [E\, S_{N-n+1}] \right\}$$

Now, replacing

$$\sum_{n=1}^{N} [S_{N-n+1}] \quad \text{and} \quad \sum_{n=1}^{N} [E\, S_{N-n+1}]$$

by equations (26) and (25a), respectively, we affect a separation of terms involving the enzyme concentration.

$$\frac{d[E]}{dt} = -k_1[E]^2 - (k_1+k_2+k_1 (S_N(0) - E(0)))[E] + (k_{-1}+k_2) E(0) \tag{27}$$

If we multiply through by -1 and let

$$a = +k_1 \qquad b = +(k_{-1}+k_2+k_1 (S_N(0) - E(0))$$

and

$$c = -(k_{-1}+k_2) E(0) ,$$

then equation (27) has the form

$$\frac{d[E]}{a[E]^2 + b[E] + C} = -dt .$$

The solution to this equation is given in Dwight.[48] Since $b^2-4ac > 0$,

$$-t = \frac{1}{\sqrt{b^2-4ac}} \log \left| \frac{2a[E] + b - \sqrt{(b^2 - 4ac)}}{2a[E] + b + \sqrt{b^2 - 4ac)}} \right| .$$

Using the boundary condition E(0) at t = 0, we have

$$\left(\frac{2a[E] + b - \sqrt{b^2-4ac}}{2a[E(0)] + b - \sqrt{b^2-4ac}} \right) \left(\frac{2a[E(0)] + b + \sqrt{b^2-4ac}}{2a[E] + b + \sqrt{b^2-4ac}} \right) = e^{-(\sqrt{b^2-4ac})t}$$

This equation shows that E decreases monotonicly with time. In the limit as $t \to \infty$, the right half of the equation must go to zero, which implies

$$2 a E (\infty) + B - \sqrt{b^2-4ac} = 0$$

or

$$E(\infty) = \frac{-b + \sqrt{b^2-4ac}}{2a} . \tag{28}$$

To evaluate this, suppose $k_{-1} = k_2 = 1$; $k_1 = 1/E(0)$, and $S_N(0) = E(0)$. This means that $a = 1/E(0)$, $b = 2$, $c = -2\,E(0)$. Therefore,

$$\frac{E(\infty)}{E(0)} = \frac{-2 + \sqrt{12}}{2} = 0.73 \ .$$

If $k_{-1} = 0$, instead, we would have $E(\infty)/E(0) = 0.62$. Thus it can be seen that in the early stages of the reaction the concentration of a free enzyme is fairly close to constant even if there is as much substrate as enzyme. We could modify the equations given earlier to take the known change of enzyme concentration into account, but this would not qualitatively change our results. In the optimal case, where $E(0) >> S_N(0)$, it is easily seen that the approximation of constant enzyme concentration approaches the exact solution.

Even under conditions which approach the steady state, the constant enzyme approximation is not a large error. If we have

$$S_N(0) = 101\,E(0), \quad k_{-1} = k_2 = 1, \quad \text{and } k_1 = 1/(100\,E(0)),$$

this will satisfy the condition that $K_m >> E(0)$. Using equation (28), we find that $E(\infty)/E(0) = 0.65$.

6. General Results of Kinetic Calculations

To obtain sequence information from a kinetic study, it is essential to have each monomeric element liberated in a time interval which is distinct from the time of appearance of other units. There is, of course, the further stipulation that a quantitative analysis must be performable in such a time interval. Under typical experimental conditions, only one nucleotide is liberated every 10 to 15 minutes.[173,17] Thus there will be no problems in taking samples from the reaction mixture. As the degra-

dation proceeds, the inevitable appearance of additional units will tend to blur the picture, and there will be practical limits beyond which no inferences concerning the sequence can be obtained. The extent to which such a technique will be feasible clearly must depend on the parameters characterizing the kinetics of the exoenzyme.

One of the major hurdles to be overcome in using an exonuclease to obtain sequence information is the ability to do an accurate quantitative analysis on the resulting mixture of mononucleotides. Much recent work has been done in this area, and as a result there appear to be at least five methods currently available which are capable of analyzing nucleotide mixtures with reasonable precision and very great sensitivity.

An automated density gradient electrophoresis of monomers followed by recording of the UV absorbance of the separated fractions has been developed by Technicon Instruments Corporation.[231] This has a sensitivity of 0.0007 mg of each nucleotide. An entire analysis takes about 10 minutes. Equally sensitive is the elution ion-exchange separation developed by Cohn and Uziel.[33] Both of these methods have the fundamental disadvantage that separation must be carried out before an analysis can be done.

Three methods have recently been developed which permit a rapid and accurate determination of the composition of an unseparated mixture of nucleotides. One of these uses linear programming to fit the observed UV spectra of the mixtures to their base composition.[170] A second method exploys a least squares fit.[124] The last takes advantage of the change in some of the bases upon strong UV irradiation to simplify the analysis of the multicomponent system.[80] Within the present limits

of UV spectrophotometers (10 lambda of 0.7 O.D. solution needed for a 1 cm pathlength microcell) these methods would permit a base composition analysis on less than 1.0 μg of total nucleotides. Recent work by Edstrom suggests that analyses of nucleotides can be performed with satisfactory accuracy on quantities as low as 0.2 mμg of sample.[49] Thus, there is no difficulty in analyzing small amounts of nucleotides, although it takes great diligence to attain an accuracy of more than 0.01 mole fraction.

Here the advantage of using polynucleotide phosphorylase as an exoenzyme becomes apparent since one can do the degradations in the presence of P_i^{32}, and thus be able to analyze extremely small amounts of product by isotopic techniques.

Another problem is the separation of small amounts of monomers from the remaining polymers. This can probably be easily accomplished by small Sephadex columns. Dialysis might also be used, but in the limit of small amounts of material some difficulty might be encountered due to adsorption of material onto the walls of the tubing or glassware. If the polymer has a sufficiently long chain length, it could be precipitated by acid.

As intuition would have suggested, the kinetic calculations show that the successive monomers are evolved in a smooth monotonic fashion. In Figure 1 are plotted the results of calculations of the time of appearance of B_1, B_3, B_2, and B_{25} for the two sets of kinetic parameters which correspond to the two extreme cases discussed above. This plot indicates that the nucleotides are liberated in a more sharply defined time interval for $\alpha_1 = \alpha_2$ than for the steady state condition. In the case that $\alpha_1 = \alpha_2$, $ES_N(0)$ has been set equal to zero. This is not

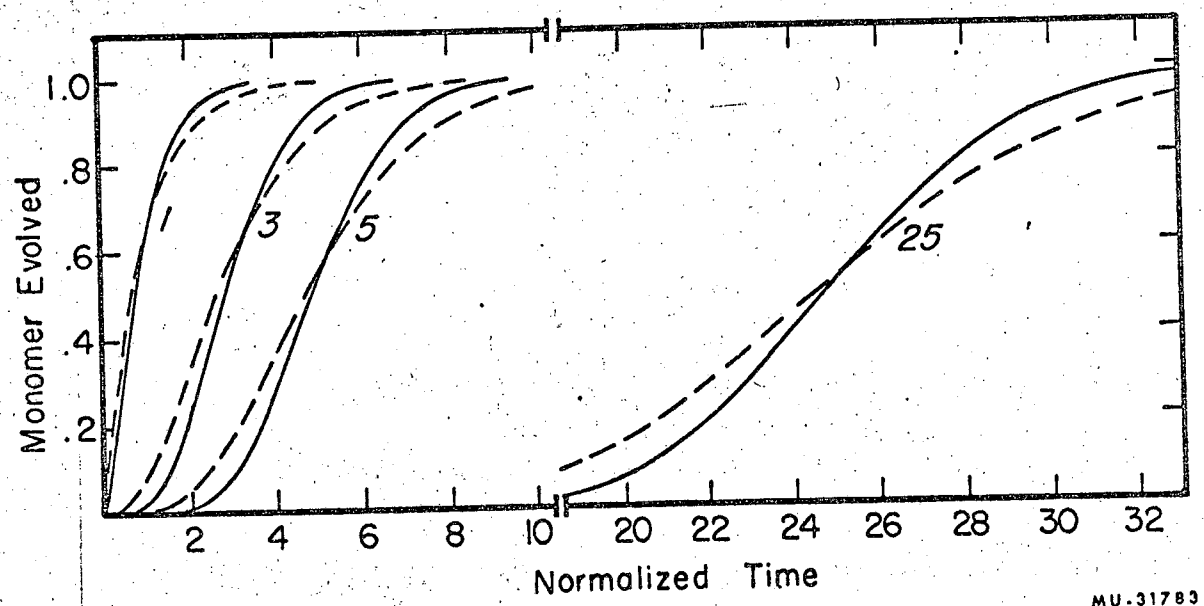


Figure 1. The liberation of monomer number n as a function of time for n = 1, 3, 5, and 25.

- - - - - steady state; $\beta = 1$, $S_N(0) = 1$

——— $\alpha_1 = \alpha_2 = 1$; $S_N(0) = 1$, $ES_N(0) = 0$

rigorously consistent with the assumed constancy of (E) unless $(E)_o >> S_N(O)$, in which event $(E) \cong (E)_o$ throughout the course of the reaction. However, further calculations reveal that assigning a non-zero value for the initial concentration of ES_N simply displaces the curves towards earlier times without altering their shapes. Consequently, the distinction between the two types of kinetic behavior is maintained.

In order to facilitate comparison of the effect of different choices of parameters, it is profitable to define a dimensionless sharpness parameter, P_n,

$$P_n = \frac{\text{time when } B_n = 0.9 - \text{time when } B_n = 0.1}{\text{time when } B_n = 0.5 - \text{time when } B_{n-1} = 0.5} \quad . \tag{29}$$

This parameter should provide a good quantitative index for the qualitative considerations outlined above. The smaller P_n, the more effective in resolution. Table III indicates that P_n for each of the choices of kinetic parameters (as calculated by equation (16) is less than that corresponding to the steady state, and greater than that for $\alpha_1 = \alpha_2$. Calculations indicate that the difference in time between $B_{n-1} = 0.5$ and $B_n = 0.5$ is approximately independent of n. Thus from equation (29), P_n must principally reflect the time of liberation of B_n.

From Figure 2 it can be seen that for both the steady state condition and for $\alpha_1 = \alpha_2$, P_n increases with increasing n. It is to be noted that the 13th monomeric unit for the steady state condition is only as well resolved as the 24th for $\alpha_1 = \alpha_2$. That the latter condition always provides better resolution can be traced to the effective absence of a dissociation step ($k_{-1} << k_1(E)$ and k_2) which would lead to an undegraded chain of intermediate size. For in this extreme case, an

TABLE III

Sharpness Parameter, P_n, for Several Sets of Kinetic Parameters

			Monomer Number (n)				
k	k_{-1}	k_2	1	2	3	4	5
Steady state			2.2	3.4	4.2	4.9	5.6
10^{-3}	1	1	2.2	3.4	4.2	5.0	5.6
10^{-4}	1	1*	2.2	3.4	4.2	5.0	5.6
10^{3}	1	1	2.2	3.4	4.2	5.0	5.6
10^{3}	1	10^{-4}	2.2	3.4	4.2	5.0	5.6
1	1	1*	2.2	3.3	3.9	4.5	4.9
10^{3}	10^{-3}	10^{2}	2.2	3.3	4.0	4.6	5.2
1.5	0.16	1.3	1.8	2.7	3.3	3.7	4.3
1100	10^{-3}	10^{3}	1.8	2.6	3.1	-	-
α_1 =	α_2		1.7	2.5	3.1	3.6	4.0

* The calculated P_n's were identical, whether $ES_N(0) = 1$ or $ES_N(0) = 0$.

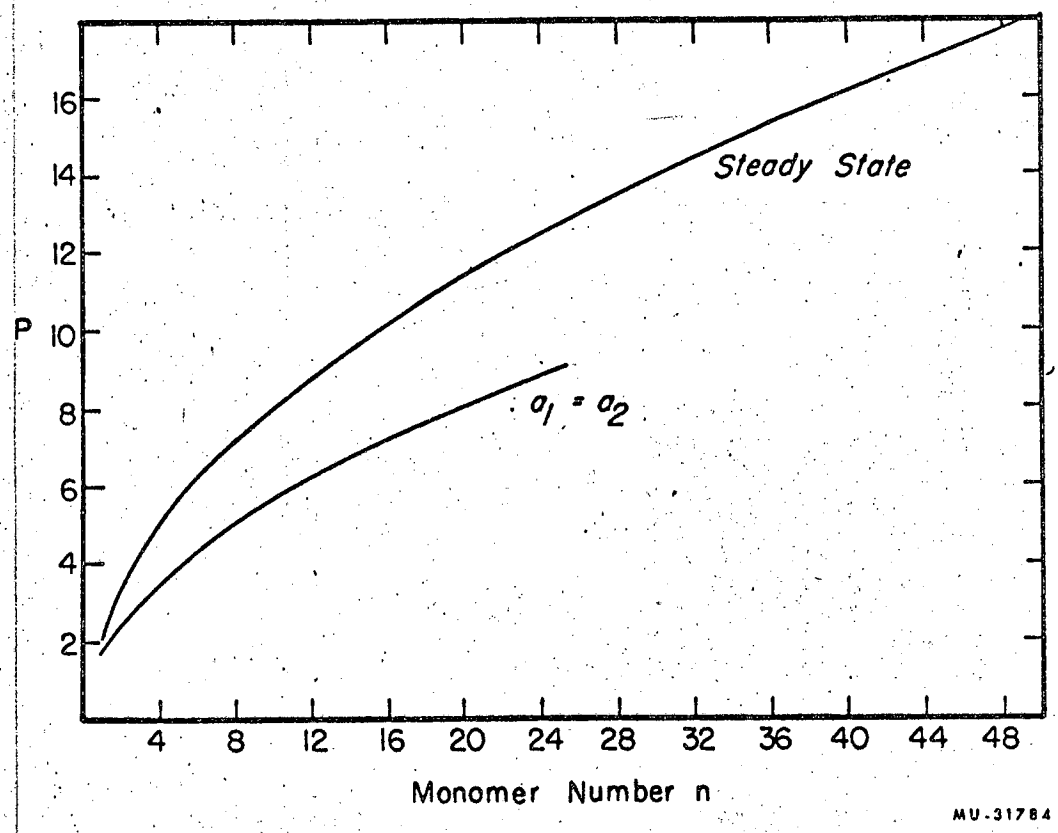


Figure 2. The sharpness parameter P_n as a function of monomer number for the steady state case and the case $\alpha_1 = \alpha_2$.

enzyme-substrate complex, once formed, can only go on to yield a chain one unit shorter.

From an experimental viewpoint, the difficulty in analyzing the mixtures of liberated nucleotides will increase as more and more nucleotides are evolved. It is highly desirable to minimize the number of units being liberated at the same time. Figure 3 illustrates a relative measure of the width of the distribution of nucleotides: the number of other units present between .1 $S_N(0)$ and .9 $S_N(0)$ when the nth unit has reached a concentration of .5 $S_N(0)$. As before, the case $\alpha_1 = \alpha_2$ provides the clearest resolution, and the steady state offers the least resolution of monomeric units.

7. Application to Sequence Studies

In order to apply the above kinetic considerations to the deduction of a nucleotide sequence, an experimental protocol must be developed. A straightforward procedure might consist of (a) a rapid mixing of exoenzyme and RNA solutions, followed by (b) removal of aliquots with rapid inactivation of the enzyme to halt the degradation, (c) removal of monomers from remaining polymers, and (d) subsequent analysis of the mononucleotides by chromatographic or spectroscopic means.

We shall discuss two concrete examples to illustrate the foregoing procedure. For the first, we take an endgroup analysis of a polynucleotide. The first four nucleotides liberated by exoenzyme attack are, for example, 2 G's and 2 A's. There are six possible arrangements of these four nucleotides. In Figures 4a and 4c, the concentrations of G and A are plotted as a function of time for two of the possible sequences,

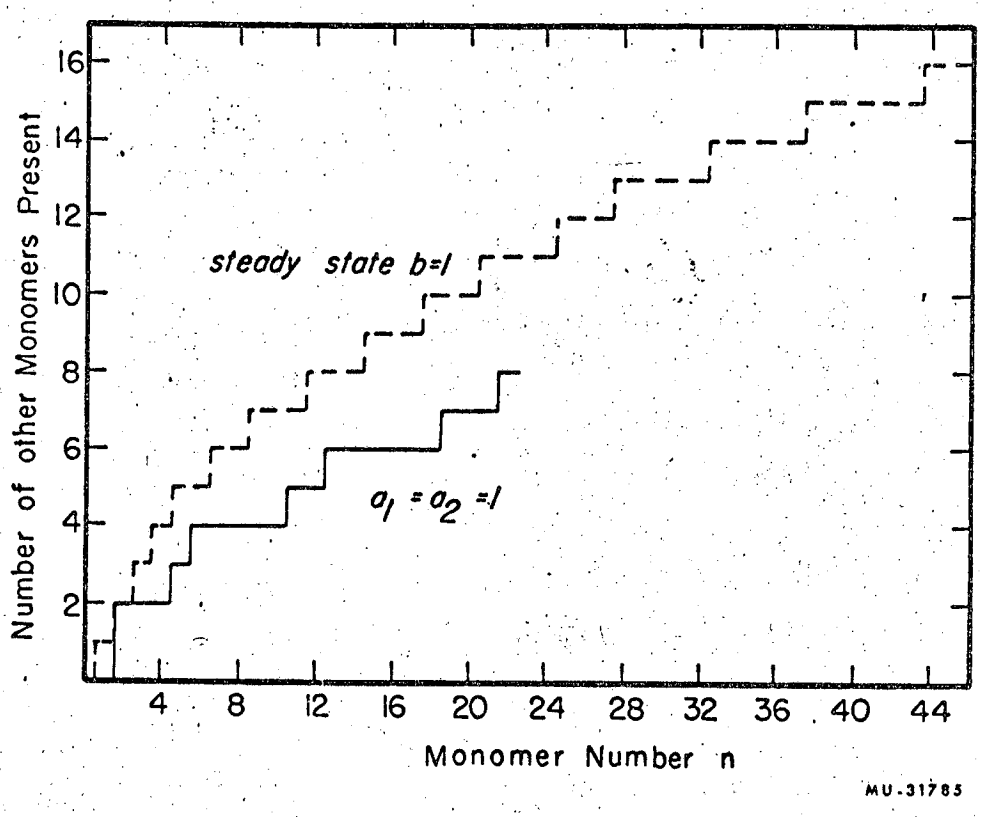


Figure 3. The number of other monomer units present between .1 $S_N(0)$ and .9 $S_N(0)$ when monomer number n reaches a concentration of .5 $S_N(0)$.

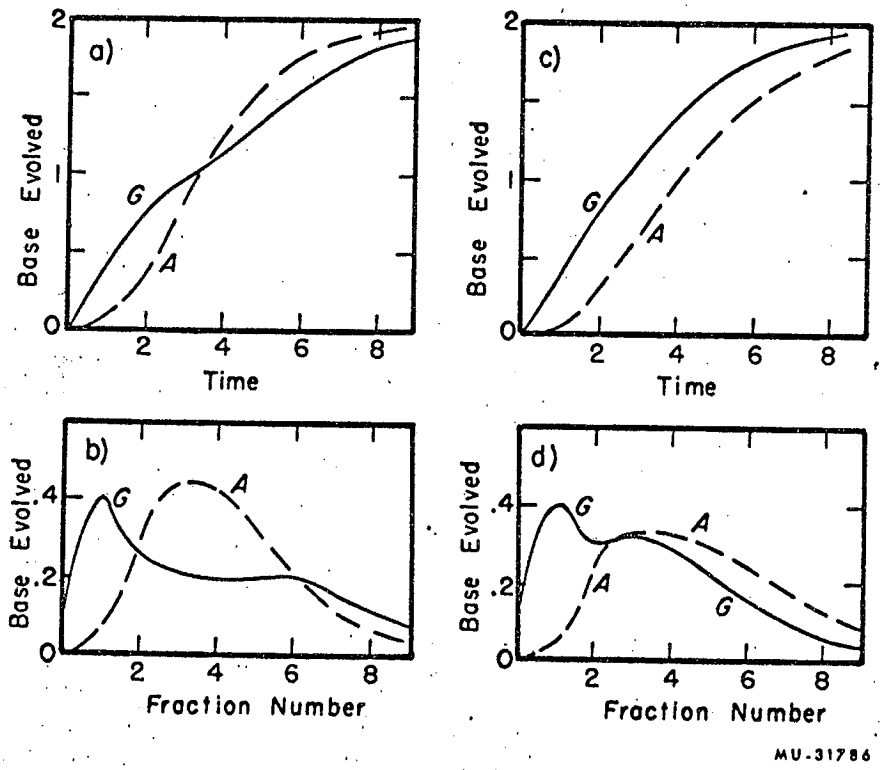


Figure 4. Two methods of distinguishing GAAG from GAGA. See text for explanation. All data calculated for $k_1 B = 1.5$; $k_{+1} = 0.16$; $k_2 = 1.3$

a) GAAG: Data from aliquots; b) GAAG: Data from fractions; c) GAGA: Data from aliquots; d) GAGA: Data from fractions.

GAAG and GAGA. These calculations have been made for the near optical kinetic conditions $k = 1.5$, $k_2 = 1.3$, $k_{-1} = 0.16$. A rather clear kinetic distinction is afforded between these two possibilities. Figures 4b and 4d indicate the results of a different method of treating the kinetic data. The quantity of mononucleotides isolated from degradation of the polymer in a given interval of time is plotted against the time of isolation, represented as fraction number as in a chromatographic analysis. Such "derivative" curves also provide a clear distinction between the two sequences.

As a second example we might consider the kinetics of exoenzyme action on a dodecamer. Figure 5 displays the calculated concentrations of the four nucleotides as a function of time under the now familiar optimal conditions. The curves have, of course, been calculated for a specific sequence of bases, but this sequence, we feel, can be rather distinctly inferred from the kinetic data as presented.

First from the asymptotic limit, it is apparent that the composition of the oligomer must be 3A, 3G, 3U, and 3C. The time course of the liberation of total mononucleotide in general permits a determination of the kinetic parameter. For the condition $\alpha_1 = \alpha_2 = \alpha$, α can be found from equation (24). Then, calculation of the curves corresponding to the carious B_n from equation (20) reveals that B_1, B_2, B_4, and B_5 coincide with, respectively, C, U, G, and A up to a concentration of about 0.8. Therefore, the indicated sequence is

C U _ G A _ _ _ _ _ _ _ .

The twelfth position must be occupied by A, as it is the last base to reach its asymptotic concentration. Further inspection of the curves

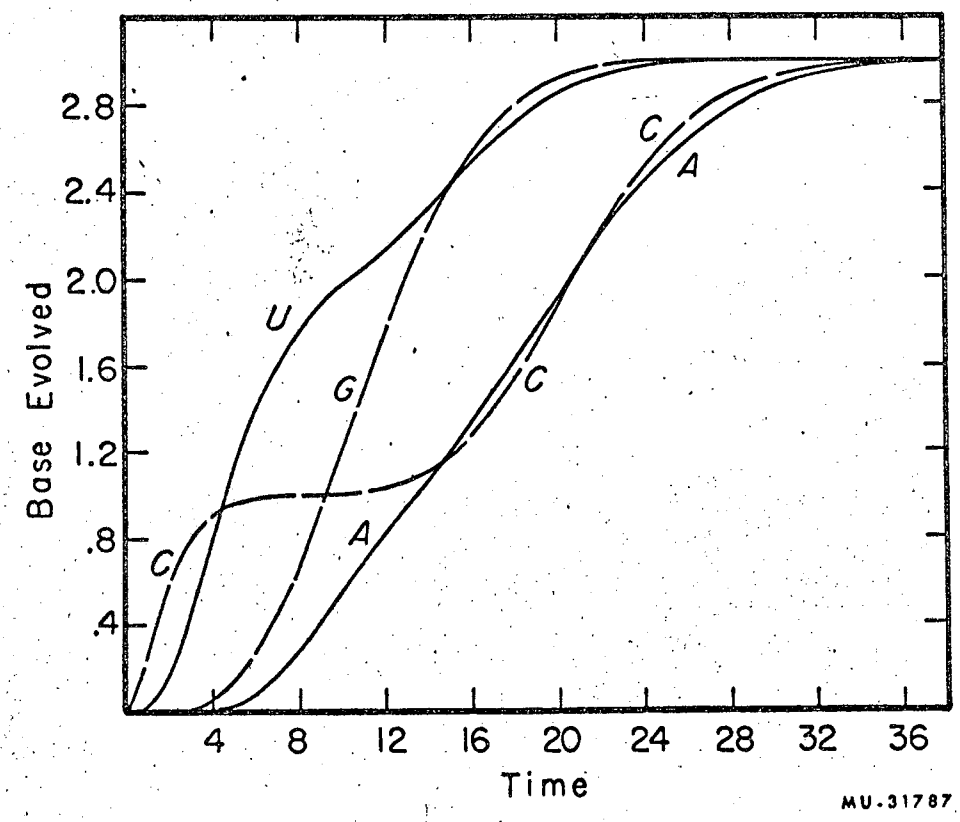


Figure 5. Predicted kinetic data from degradation of CUUGAGGUACCA by an exonuclease: $\alpha_1 = \alpha_2 = 1$.

suggests that the third nucleotide is U, as its concentration exceeds unity before the fourth nomonucleotide (G) has begun to appear. The known sequence is now

C U U G A _ _ _ _ _ _ _ A.

It should be noted from equation (20) that $\ddot{B}_n = 0$ when $t = (2n-1)/\alpha$ ($t = (n-1)/\beta$ for the steady state condition). A shoulder appears in the curve for the evolution of U at $t = 15$. Since α has been taken equal to unity position 8 must belong to U. All three U's have been accounted for, and from the curve for G it is evident that all three G's appear before all the U is liberated. Hence the 6th and 7th positions in the sequence must be occupied by G. This leads to the partially completed sequence

C U U G A G G U _ _ _ A.

Positions 9, 10, and 11 must be filled by 2 C's and one A. There are three possibilities, (a) -CCAA, (b) -ACCA, and (c) -CACA. (The terminal position on the right is already known to be occupied by A.) Focusing attention on the curves for the liberation of A and C, one sees that the former intersects the latter. The concentration of evolved A exceeds that of C at a stage in the reaction, only to fall below it again at a later time. Possibilities (a) and (c) would always require the concentration of C to exceed that of A. With possibility (b) as the only remaining one, the inferred sequence must be

C U U G A G G U A C C A .

In actual practice, a hypothetical sequence must be carefully checked by the comparison of the computed curves suggested by it with those obtained

experimentally over the entire course of the reaction for all four mononucleotides. The last step in the above example involved a small difference in the liberation of C and A at a late stage in the reaction. An experimental distinction, unlike the one proferred by these calculations may be virtually impossible due to unavoidable imprecisions in the measurements. Here the practical limitations inherent in any procedure of this type my intrude.

By contrast, the steady state kinetic curves for the liberation of the four mononucleotides, shown in Figure 6, fail to exhibit some of the features present in Figure 5. In particular, the shoulder in the curve for U, the intersection of the A and C curves, and the details of the four curves near the asymptotes are virtually completely obscured.

From the preceding discussion it appears that the efficiency of an enzymatic approach to sequence studies can be greatly enhanced by adjustment of experimental conditions away from the steady state situation. The use of high enzyme concentrations, of the order of the Michaelis constant, where possible, is recommended.

8. Limits of Resolution

If we can adjust the kinetic parameters experimentally, it will be possible to determine the sequence many bases in from the end of a polynucleotide. Given the accuracy of our ability to analyze the base composition of the resulting monomer mixture we could, in principle, determine the maximum length of RNA whose sequence is determinable by the exoenzyme method. Conversely, if we wanted to determine the sequence of an n-mer, we would know what accuracy of analysis is needed. To do this we would have to calculate the kinetics of all 4^n base sequences.

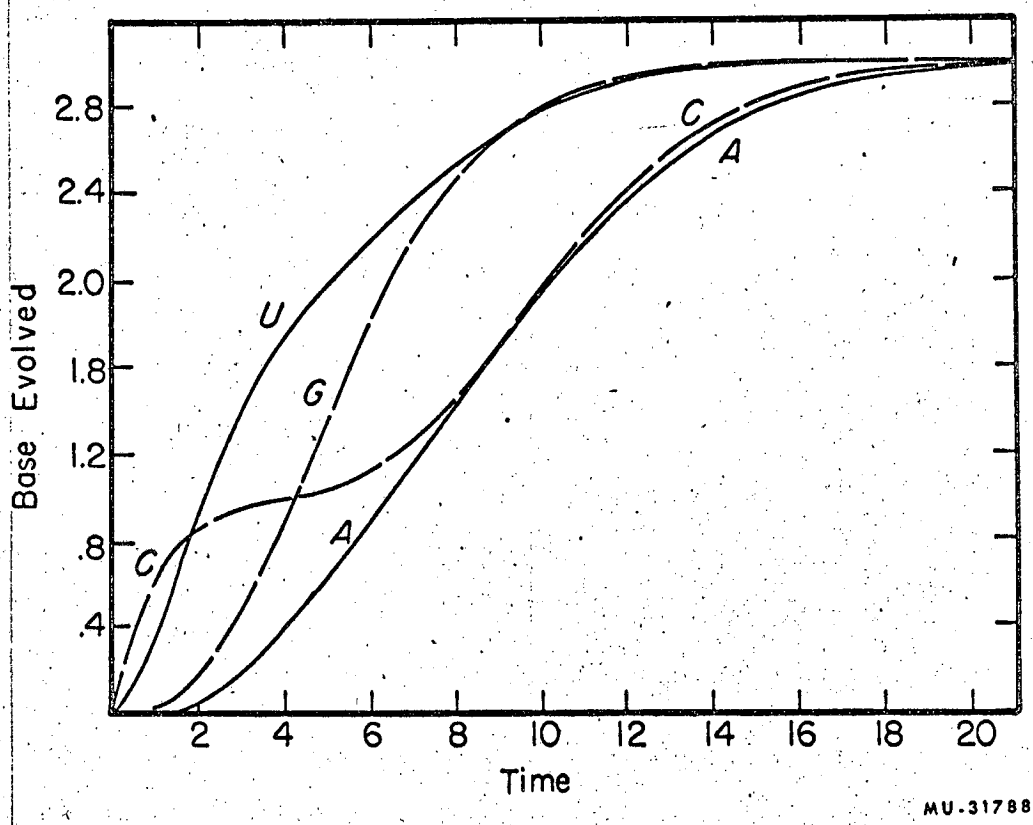


Figure 6. Predicted kinetic data from degradation of CUUGAGGUACCA by an exonuclease: steady state conditions: $\beta = 1$.

We would then compare each of these with all the others--an impossible total of $(4^n)(4^n-1)/2$ comparisons.

Instead, we will make some estimates of the resolvability of different sequences by comparing the kinetics of appearance of several sequence isomers as a function of how far in from the end they are placed. We consider the six sequence isomers: AABB, ABAB, ABBA, BAAB, BABA, BBAA. These were chosen because they are moderately difficult to resolve. In order to further simplify the problem, it was assumed that there are no A's or B's adjacent to this sequence. The last assumption we make is that the base composition of the fragment analyzed is random; it contains equal amounts of A, B, C, and D. Thus, if AABB occurs in positions 16 to 20 of the polymer, random base composition implies that three A's and three B's appeared in the reaction mixture previously.

What we wish to know is the maximum difference between the amounts of A or B evolved from the different sequence isomers at a given time, and the total concentration of A or B in the reaction mixture at that time. It turns out that it is only necessary to consider nine of the 15 possible comparisons among the sequence isomers with composition A_2B_2. For example, if we can distinguish ABBA from ABAB, we can also distinguish BAAB from BABA, since we have just interchanged A and B. The most effective times for comparing the monomer composition evolved from two different sequences are the times at which the rate of evolution of each successive monomer is a maximum. For the optimal case this is $T = (2n-1)/\alpha$ where n is the position in the chain of each of the four monomers under consideration. We calculated the monomer compositions at these four times for five choices of the position of the six sequence isomers. We let each isomer appear in positions 1 to 4, 5 to 8, 9 to 12,

13 to 16, and 17 to 20. The results of these comparisons are summarized in Table IV. The analysis is based on the monomer which appears nearest to the end of the polymer degraded by the enzyme. In the cases we have considered this is always A. Max ΔC is the difference in concentration of A evolved from the two sequence isomers compared at the time shown. This was the maximum difference found. Average C is the expected concentration of A released at the time shown for a polymer of random base composition.

Several conclusions can be drawn from Table IV. The most difficult sequences to resolve are ABBA and BAAB. The resolution between any pair of sequences gets worse as the number of residues preceeding the sequence increases. Max ΔC/Average C is essentially the accuracy to which we must be able to determine base composition in order to decide the sequence. Thus, 5% accuracy is sufficient to sequence an average 4-, 8-, or 12-mer. To extend the analysis to 16 bases we would need 2% accuracy, and 1% accuracy is the minimum necessary to make an attempt at sequencing a 20-mer worthwhile. Since enzymes are available which attack nucleic acids from both the 3' and 5' ends, the use of two different exonucleases may permit the sequence determination of nucleic acid fragments of up to 40 nucleotides in length, if we have 1% accuracy in analysis.

9. Application to Experimental Data

The one natural RNA in which limited exonucleolytic digestion has been performed in an attempt to learn the terminal sequence is TMV-RNA. This RNA is one of the longest known, and thus great care must be taken

TABLE IV

Resolution of Sequence Isomers

Sequences Compared		Position	Time/α	Maximum ΔC	Max ΔC/Average C
AABB	BBAA	1-4	5	1.078	.635
AABB	ABAB	1-4	5	.351	.207
AABB	BABA	1-4	5	.827	.488
AABB	ABBA	1-4	5	.602	.356
ABAB	BABA	1-4	3	.520	.586
ABAB	ABBA	1-4	7	.298	.176
ABBA	BABA	1-4	3	.448	.550
ABBA	BBAA	1-4	3	.717	.883
ABBA	BAAB	1-4	3	.376	.463
AABB	BBAA	5-8	13	.818	.551
AABB	ABAB	5-8	13	.220	.150
AABB	BABA	5-8	13	.798	.539
AABB	ABBA	5-8	13	.411	.276
ABAB	BABA	5-8	13	.378	.298
ABAB	ABBA	5-8	15	.205	.130
ABBA	BABA	5-8	11	.239	.318
ABBA	BBAA	5-8	11	.441	.584
ABBA	BAAB	5-8	9	.164	.378
AABB	BBAA	9-12	21	.663	.255
AABB	ABAB	9-12	21	.174	.073
AABB	BABA	9-12	21	.489	.204
AABB	ABBA	9-12	21	.332	.139
ABAB	BABA	9-12	21	.315	.142

Sequences Compared		Position	Time/α	Maximum ΔC	Max ΔC/Average C
AABA	ABBA	9-12	23	.166	.067
ABBA	BABA	9-12	19	.183	.103
ABBA	BBAA	9-12	19	.347	.196
ABBA	BAAB	9-12	17	.096	.064
AABB	BBAA	13-16	29	.571	.171
AABB	ABAB	13-16	29	.148	.045
AABB	BABA	13-16	29	.423	.126
AABB	ABBA	13-16	29	.286	.086
ABAB	BABA	13-16	29	.275	.084
ABAB	ABBA	13-16	31	.143	.042
ABBA	BABA	13-16	27	.153	.055
ABBA	BBAA	13-16	27	.295	.106
ABBA	BAAB	13-16	25	.065	.025
AABB	BBAA	17-20	37	.510	.118
AABB	ABAB	17-20	37	.131	.030
AABB	BABA	17-20	37	.379	.088
AABB	ABBA	17-20	37	.255	.059
ABAB	BABA	17-20	37	.248	.059
ABAB	ABBA	17-20	39	.127	.029
ABBA	BABA	17-20	35	.135	.036
ABBA	BBAA	17-20	35	.262	.069
ABBA	BAAB	17-20	33	.048	.013

in its preparation lest the molecule by fractured by nucleases, heating, or other rough treatment. Currently the best preparations of TMV-RNA are estimated by sedimentation studies to be 80% unbroken.[143] Since TMV-RNA has over 6000 base residues, the problem of detecting just one of them becomes formidable. Careful work by Sugiyama and Fraenkel-Conrat[228] and by Whitfeld[257] has led to the conclusion that the 3' terminal nucleoside (5' linked) is A. Both of these methods involved degradation of the nucleic acid and isolation, by isotope dilution, of the one nucleoside which must have come from the 3' end. The yields were never very good. This may be due to phosphate exchange between the terminal nucleoside and nucleoside cyclic phosphates which are present during the degradation.[122] Thus, with the terminal nucleoside already known, TMV-RNA became a prime candidate for the use of an exonuclease, and several studies have been carried out by Fraenkel-Conrat and his collaborators.

Venom phosphodiesterase was the first enzyme tried by Singer and Fraenkel-Conrat,[205] since it is readily available and is known to have a very low degree of contamination with endonuclease. Using the assumption that the molecular weight of the enzyme is 20,000, one can estimate that the ratio of moles of enzyme used to moles of substrate was around 0.1 for most experiments. The concentration of RNA was 0.2 - 1.0 mg RNA in a total volume of 0.25 ml. This means that the average enzyme concentration used was about 10^{-7} molar. K_m for various substrates of venom PDE is 10^{-4} to 10^{-3} molar.[177] Thus, under these conditions $K_m >> E(0)$. Thus, the steady state conditions should be used for calculating the kinetics.

The less than ideal experimental protocol used by Singer and Fraenkel-Conrat was necessitated by the difficulties of working with such small amounts of material. Digestion was performed at 0°C with the above mentioned concentrations of C^{14}-RNA and enzyme. 0.1 M NaCl and 0.004 M $MgCl_2$ were added to promote structure formation of the RNA. This seemed to have an inhibitory effect on the random release of nucleotides, which was observed either at higher temperatures or in the absence of these salts, and was attributed to endoenzyme activity. The buffer was 0.004 M pH 8.8 borate. After 15 minutes of reaction 100 µg of each of the four unlabeled nucleotides was added. The RNA was precipitated with EtOH, and the nucleotides were adsorbed from the alcohol supernatant by charcoal, separated by electrophoresis and counted. The RNA was resuspended in buffer and treated with two more batches of enzyme in the above manner. This protocol leaves much to be desired, kinetically, since each readdition will cause needless loss in sequence information due to the time lag which occurs in the reformation of the various ES_{N-n+1}. But it permits less RNA to be used, since all of the monomers released after each time interval can be isolated.

The data of Singer and Fraenkel-Conrat[205] for the total amount of each nucleotide evolved up through and including each of the three additions of enzyme is summarized below for the average of two runs. Each step represents 15 minutes of hydrolysis.

Step	A	G	U	C	Total
1.	0.90	0.14	0.29	0.22	1.54
2.	1.32	0.27	0.68	0.47	2.72
3.	1.78	0.44	1.14	0.72	4.10

From these data the authors concluded that there are two possible sequences:

--UpApCpUpA or --UpApUpCpA

Later work in Fraenkel-Conrat's laboratory with polynucleotide phosphorylase as the exoenzyme led to the proposal of a third possible sequence,[64]

--UpApGpCpUpA

Given sufficiently accurate kinetic data, it should be possible to distinguish among these three possibilities, but our analysis of the data presented above is disappointing in this respect. In Figure 7, the experimental data of Singer and Fraenkel-Conrat is compared with that calculated for the three proposed sequences. The agreement is none too good for each of the three, although --UpApCpUpA seems to fit the qualitative behavior of the experimental curves much better than the other two possibilities. A real difficulty here may be the presence of broken ends leading to spurious evolution of nucleotides--but in order to correct for this, we would have to know the homogeneity of the particular TMV-RNA preparation used in these experiments, and the base specificity (if any) of the breaks in the RNA.

A disturbing factor is the recent work of Steinschneider and Fraenkel-Conrat,[225] which proposes still another sequence as a result of terminal chemical degradation:

--GpCpCpCpA.

This is obviously in almost complete disagreement with the previously published exoenzyme work. It would be impossible to fit the venom PDE kinetic data thus far obtained to this sequence. Thus the terminal se-

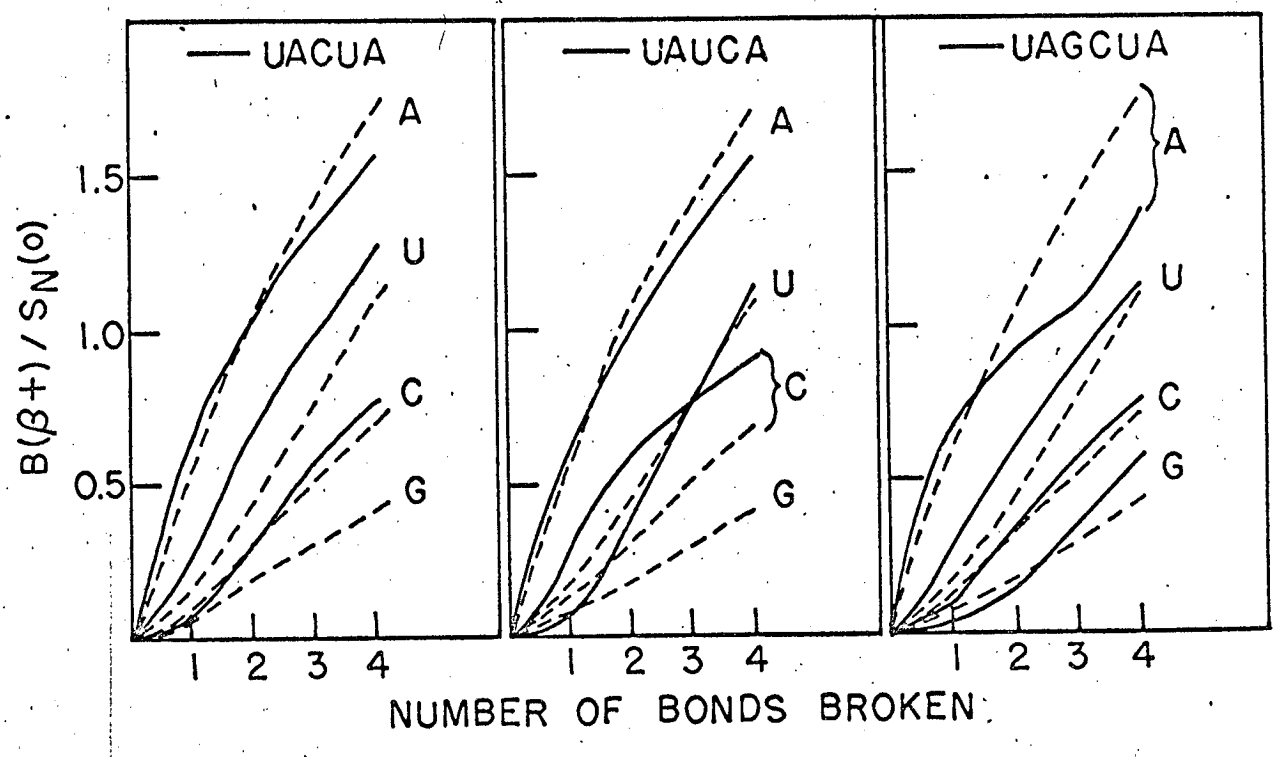


Figure 7. Degradation of TMV-RNA by venom phosphodiesterase.

- - - - - - - kinetic data of Singer and Fraenkel-Conrat.[205]

——————— calculated results for proposed sequence.

quence of TMV-RNA remains well studied, but unproven. Recent work by Mandeles suggests some evidence for a G in the fourth position from the end, but this awaits further confirmation.[143]

Several experiments on the kinetics of degradation of small oligonucleotides of known sequence have been published. These are done to prove the exo nature of the enzyme, but they give some insight as to whether or not the enzyme follows anything resembling the kinetic scheme we have discussed. Fiers and Khorana[60] studied the kinetics of degradation of ApApApA by L. acidophilus phosphodiesterase. They found that the fragments were liberated in the expected order, but that the maximum concentration reached by the intermediate ApApA is less than would be expected if the attack were random. The data is reproduced in Figure 8a. At all times, the concentration of degradable intermediates is less than the concentration of undegraded starting materials. This may reflect the fact that these enzymes prefer to degrade shorter species. But this seems to be in conflict with the results reported in the same paper that TpTpTpT, TpTpT and TpT are all hydrolyzed at roughly the same overall rate. Another possible explanation is that the enzyme once attached to the polymer chews off several monomers at once in rapid order before releasing the substrate. The implications of such a kinetic scheme will be discussed later.

In contrast to these results are the earlier experiments by Razzel and Khorana[178] on the kinetics of the degradation of TpTpTpTpT by venom PDE. These results, reproduced in Figure 8b, show that in this case the concentration of undegraded TpTpTpTpT falls to about 1/e at about the same time as the concentration of TpTpTpT reaches its maximum. The two curves then cross. The maximum concentrations reached by TpTpTpT

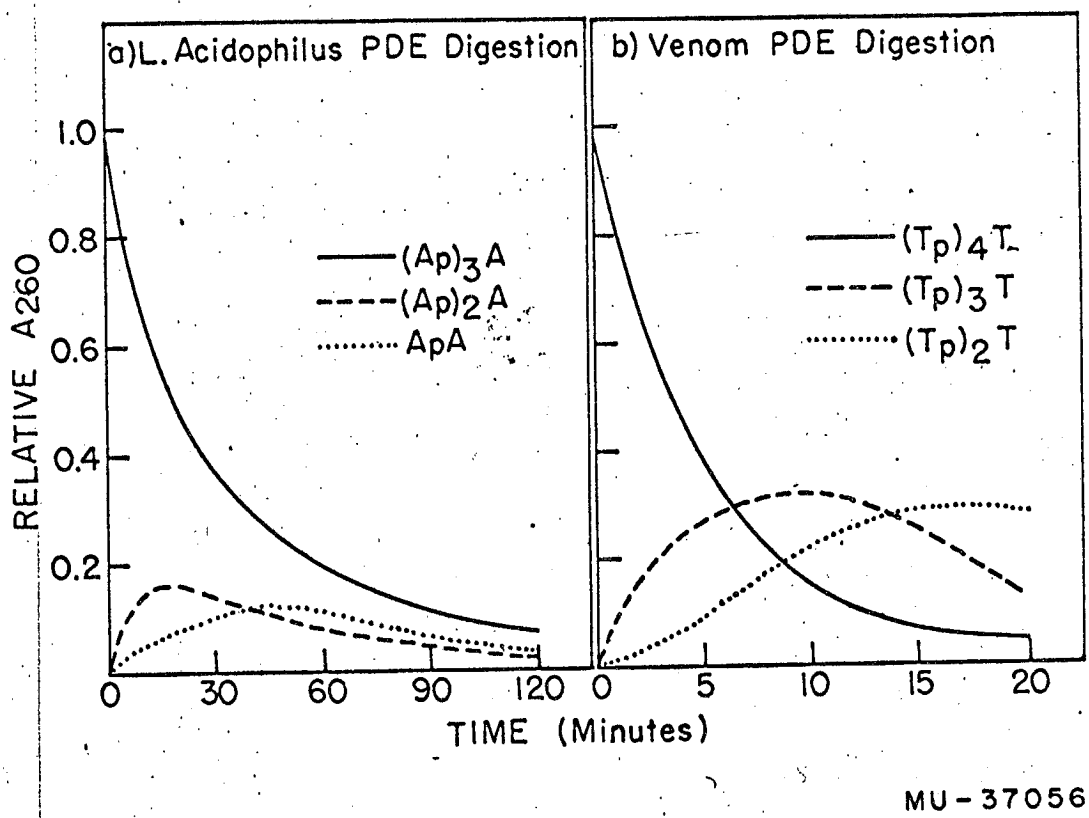


Figure 8. Degradation of oligonucleotides by exonucleases.

a) L. acidophilus phosphodiesterase degradation of $(Ap)_3A$.[60]

b) Venom phosphodiesterase degradation of $(Tp)_4T$.[178]

and TpTpT are both about .31 compared with values calculated from equation (21) of 0.37 and 0.27. These observations are consistent with the steady state mode of the random model we have discussed.

The work of Nihei and Cantoni[162] also confirms that venom PDE acts according to the random model of attack. They show that the distribution of S_{N-n}'s is approximately Poissonian, as would be predicted by our model. But their substrate, mixed yeast s-RNA, is much too complicated to permit a detailed analysis of their kinetic data. They are able to conclude only that the distribution of many of the various nucleotides is non-random, and that the nucleotides methyl A, methyl G, and ψ are located near the center of the RNA.

10. Effect of Endoenzyme Impurities

In the previous discussion we have made frequent reference to the necessity of using an exoenzyme with a low contamination of endonuclease activity. Now that we have a fairly satisfactory model for the kinetics of exoenzyme attack we can place some upper limits on the endoenzyme impurities that can be tolerated. These limits will, of course, depend strongly on the use to which the enzyme is put. The experimental quantity of interest is the ratio of endo to exo activity, either for a mixture of enzymes or for a pure, but only partially specific enzyme like micrococcal nuclease (EC 3.1.4.7)[150,258]. This ratio, which we shall abbreviate as R, is the number of bonds broken per unit time in an endonucleolytic mode divided by the number of bonds broken by exonucleolytic attack. To simplify the problem, we shall limit our attention to times in which the fraction of total bonds broken is small, and to polymers of

long chain length. We shall also assume that concentration of enzyme is high enough or concentration of substrate is low enough to avoid saturation of the enzyme.

In the steady state case for an exoenzyme, the total rate of monomer release is constant.

$$\sum_{n=1}^{N} \frac{dB_n}{dt} = \beta S_N(0)$$

Suppose that the endoenzyme impurity breaks bonds randomly at the rate of M per unit time. Within the above approximations M is a constnat, depending only on the initial substrate concentration. For each bond broken it exposes an end susceptible to exonuclease attack. Such a case might be an impurity with the characteristics of pork liver nuclease[229] in venom PDE, or a T2 type ribonuclease[51] impurity in L. acidophilus PDE. If the endo impurity has no base specificity, the broken ends will have random base composition. Thus monomers released from these ends due to sufficient exonuclease action will just be a random mixture corresponding to the base composition of the polymer.

The ratio of the number of broken ends to the number of initial ends is just $Mt/S_N(0)$. The total rate of monomer release now becomes

$$\sum_{n} \frac{dB_n}{dt} = \beta S_N(0) + \beta Mt,$$

where the assumption has been made that the endoenzyme itself releases no monomer. This assumption is valid if the polymer is very long. Thus we can see that a small endo impurity will eventually swamp out the non random distribution of bases liberated from the true end of the polymer.

The concentration of monomers in solution is, for small times

$$\sum_n B_n = \beta S_N(0)\, t + \frac{\beta M t^2}{2} \tag{30}$$

The relative endo and exo activities was previously defined.

$$R = \frac{M}{\beta S_N(0)} \tag{31}$$

Therefore, the relative concentration of monomers released from broken ends to true ends can be written as

$$\frac{\text{noise}}{\text{signal}} = \frac{\beta M t^2/2}{\beta t S_N(0)} = \frac{Mt}{2S_N(0)} = \beta t \frac{R}{2} \tag{32}$$

Using the results shown in Table II, we can calculate the purity, R, required to determine the sequences of oligomers of various lengths. To do this we take advantage of the fact that βt is equal to the number of monomers released in time t.

$$\frac{\text{noise}}{\text{signal}} = \frac{nR}{2} \tag{33}$$

Several values of n and R are given below for the maximum permittable noise/signal useful for sequencing.

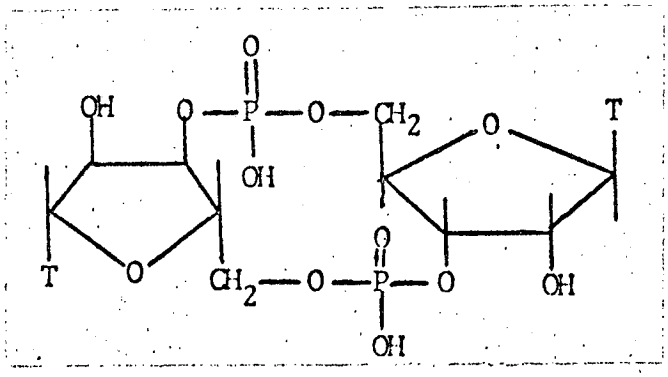
$\frac{\text{noise}}{\text{signal}}$	n	R
.05	5	.02
.02	10	.004
.01	20	.001

Since the endo impurity in L. acidophilus PDE is estimated as less than 0.1% of the activity,[60] this enzyme is useful for sequence studies of up to the 20 terminal residues of RNA.

The presence of endonuclease impurity which is specific for certain linkages would be a serious problem. For example, if L. acidophilus PDE were contaminated with T1 ribonuclease, the noise/signal of G would rapidly become intolerable high. Even more serious problems would result if the endo impurity was conformation specific. In this case it might cleave in only a few specific places, and subsequent exo degradation of these ends would yield a non-random pattern of monomer evolution. If an exoenzyme is known to have an appreciable endo contaminant, the lack of specificity of this impurity should be carefully proven before using the enzyme to determine sequence. It should be noted that Singer and Fraenkel-Conrat have found that subjecting TMV-RNA to some unfavorable conditions leads to preferential breakage of CpA and UpA bonds to give Cp and Up.[206] It is apparent that the amount of endo impurity is a critical factor in deciding which exoenzyme to use in a sequence study. We shall next consider some of the methods which can be used to assay small endo impurities in an exoenzyme.

11. Determination of Endoenzyme Impurities

There are three types of methods we can use to detect and quantitate the presence of small endo impurities in an exoenzyme. The most obvious involves choosing a substrate which only one type of enzyme can degrade. Many such substrates are available, but few are ideal. Cyclo pTpT



and the analogous cyclo trimer and tetramer are all degraded very slowly by VPDE.[177,178] This indicates the presence of endo impurity. But it is not apparent how to relate the activity of an endoenzyme of unknown properties towards these unusual substrates with its activity towards a long linear single strand RNA. Thus, small cyclo nucleotides are useful only in qualitatively demonstrating the presence of an endo contaminant. In principle, one could make much larger cyclonucleotides by either of the two procedures shown below, where N > p means 2'3'-cyclic phosphate.

$$(Ap)_nU > p \xrightarrow[\text{very dilute}]{\text{Pancreatic RNAase}} \text{cyclo } (Ap)_nUp$$

$$(Ap)_nG > p \xrightarrow[\text{very dilute}]{\text{Ti RNAase}} \text{cyclo } (Ap)_nGp$$

These larger substrates would presumably be degraded by an endoenzyme with the same kinetics as a linear polymer. In a similar way ϕX 174 DNA[62,65] is an ideal substrate to assay the presence of DNAase impurity in VPDE. To determine the endo activity in any of the above cases, one can assume that once one bond of the cyclo nucleotide is broken the resulting linear polynucleotide is immediately degraded to monomer by the much larger activity of the exoenzyme present.

The extreme substrate specificity of some of the common exonucleases suggests other simple methods of analysis using substrates with 5' terminal phosphates or other 5' derivatives which will greatly magnify the presence of an endo impurity in L. acidophilus PDE. It is this way that Khorana was able to get an estimate of the upper limit of the endo contamination of this enzyme.[60]

A second type of method for detecting endo impurities is to analyze the mixture of products found after the enzyme has gone to completion. Nihei and Cantoni have discussed this method in detail,[162] and we will reproduce some of their argument here. Consider the nucleic acid pNp....pYpZp...pM. Hydrolysis with a 5' PDE yields only nucleotides and pNp. Base hydrolysis gives the nucleoside M as well as pNp. This permits the chain length of the substrate to be determined. Now, if there is an endonuclease impurity present it can break the bond between Y and Z. There are two possibilities. If the reaction is pN...YpZ...pM → pN...Yp + Z...pM, most 5' exoenzymes will no longer be able to degrade the chain ending in a 3' phosphate. Hydrolysis will eventually yield Z as a nucleoside, and either pN...Yp or pYp will accumulate in the solution. Thus an analysis of the products will easily present the endo activity to be estimated. If, on the other hand, the endonuclease catalyzes the reaction pN...YpZ...pM → pN...Y + pZ...pM, a different analysis is needed. The reaction is stopped after partial hydrolysis, and the reaction mixture is subjected to alkaline hydrolysis. This will yield one nucleoside for each chain ending in Y. Thus, by performing this experiment as a function of time the endo impurity can be detected. Both of these chemical methods, however, require an accurate method of analysis to detect small quantities of impurities.

The last and most sensitive method is based on the fact that a very small endo impurity will radically change the distribution of chain lengths present in the reaction mixture after partial hydrolysis has occured. While our discussion will center on the use of viscometry for detecting this, chromatography and sedimentation have also been used with relative success.

RNA under appropriate conditions behaves like a random coil.[217] Under these conditions the intrinsic viscosity is given by the well known equation:

$$[\eta] = k M^{1/2} .$$

In practice, one would first want to verify this equation for the particular substrate being used. If we have a distribution of molecular sizes we have instead

$$[\eta] = k \bar{M}_v^{1/2} \tag{34}$$

where $\bar{M}_v^{1/2}$, the viscosity average molecular weight, is in this case defined by the equation[230]

$$\bar{M}_v^{1/2} = \frac{\sum_i M_i^{3/2} N_i}{\sum_i M_i N_i} \tag{35}$$

where N_i is the concentration of species with molecular weight M_1. Using an average residue weight M_o, we can rewrite this as

$$[\eta] = k M_o^{1/2} \bar{x}_v^{1/2} .$$

$\bar{x}_v$ is the viscosity average chain length.

For exoenzyme degradation, the number of bonds broken per unit substrate is βt, and the fraction of bonds broken is

$$f_b = \frac{\beta t}{N-1} \simeq \frac{\beta t}{N} \qquad (36)$$

where N is the chain length of the undegraded polymer. The molecular weight of the substrates in solution is a very narrow distribution centered about the original chain length minus the average number of bonds broken.

$$\bar{M}_V = NM_o\,(1-f_b) \qquad (37)$$

and

$$[\eta]_{exo} = k(M_o N(1-f_b))^{1/2} = k(M_o N(1-\frac{\beta t}{N}))^{1/2}. \qquad (38)$$

For an endoenzyme, the rate of breaking bonds is proportional to the number of bonds in solution.

$$\frac{d(N\,S_N(O)}{dt} = k\,(N\,S_N(O)) \qquad (39)$$

The fraction of bonds broken is thus

$$f_b = 1-e^{-kt}\,.$$

In contrast to exo degradation, the distribution of substrates from endonucleolytic degradation of a long polymer is very broad. This distribution function is given in reference 230, p. 615. If the polymer at time zero has a chain length N, it is shown that for random endo degradation

$$[M_x] = [M_N]_o\, e^{-k(x-1)t}\,(1-e^{-kt})\left\{2 + (n-x-1)(1-e^{-kt})\right\} \qquad (40)$$

where $[M_x]$ is the concentration of x-mer at time t, and $[M_N]_o$ is the initial concentration of polymer. What we must calculate is

$$\bar{x}_v^{1/2} = \frac{\sum_x [M_x]\, x^{3/2}}{N\, S_N(0)} \quad . \tag{41}$$

But we have been unable to find a satisfactory closed form for the above finite sum. Instead, we will take advantage of the fact that $\bar{x}_n$ and $\bar{x}_w$, respectively, the number average and weight average chain length, have been computed for a mixture arising from a random degradation of a polymer. By adapting the equations in pages 616 and 617 of reference 230 to our notation, we find:

$$\bar{x}_n = \frac{\sum_x x[M_x]}{\sum_x [M_x]} = (1/N + f_b)^{-1} = N/(1 + Nf_b) \tag{42a}$$

$$\bar{x}_w = \frac{\sum_x x^2[M_x]}{\sum_x x[Mt]}$$

$$= \frac{Nf_b^2 + 2(1-f_b)\left\{(f-f_b)^N + Nf_b - 1\right\}}{Nf_b^2} \tag{42b}$$

It is known that $\bar{x}_n \leq \bar{x}_v \leq \bar{x}_w$. Thus we will be able to calculate the upper and lower bounds of $\bar{x}_v$ even though we cannot compute $\bar{x}_v$ directly. Expressions for the limits of the viscosity of a mixture arising from random endo degradation are shown below.

$$[\eta]_{max} = kM_o^{1/2}\, \bar{x}_w^{1/2} = k'\bar{x}_w^{1/2} \tag{43a}$$

$$[\eta]_{min} = kM_o^{1/2}\, \bar{x}_n^{1/2} = k'\bar{x}_n^{1/2} \tag{43b}$$

It is most convenient to plot the fractional change in $[\eta]$ as a function

of the fraction of total bonds broken. Calculated values of $[\eta]/[\eta(0)]$ are shown for three cases (a, b, and c) in Figure 9. It can be seen that the viscosity drops sharply for endo attack, calculated using either $\bar{x}_w$ (curve b) or $\bar{x}_n$ (curve c), while for exo attack (curve a) the change in viscosity is linear and very slight. To demonstrate the sensitivity of this method, the curves for an exoenzyme containing a 1% endo impurity are also shown as calculated by both $\bar{x}_w$ (curve d) and $\bar{x}_n$ (curve e). Here the difference between the two calculations is considerable. Thus we would really need a better estimate $\bar{x}_v$. But whichever curve one uses it is apparent that an endo impurity of 1% is easy to detect by this method.

Experiments similar to these calculations have been reported by Williams, Sung, and Laskowski, using venom PDE as the enzyme and DNA as the substrate.[258] Their work shows that viscometry is a simple and powerful tool for determining the nature of action of enzymes which depolymerize DNA.

12. Tight Binding Model--High Enzyme Approximation

We now want to consider an alternative model for the mechanism of exonuclease digestion of RNA. This model corresponds to the irreversible degradation of a polymer chain in the limit of strong binding of the substrate to the enzyme. In this case, once an enzyme-substrate complex is formed the whole polymer is digested one monomer unit at a time, according to the following reaction scheme.

$$E + S_N \xrightarrow{k_1} ES_N \xrightarrow{k_2} ES_{N-1} \xrightarrow{k_2} ES_{N-2}$$

$$+ B_1 \qquad + B_2$$

$$\cdots \xrightarrow{k_2} ES_1 \xrightarrow{k_3} E + B_N$$

$$+ B_{N-1}$$

There are several pieces of evidence which suggest such a mode of reaction may take place under certain conditions. This idea was first suggested by Dekker.[39] He reported evidence that when wheat germ RNA is digested by an exoenzyme no intermediate chain length RNA is observed. Only long polymer and monomer were found, even after much of the substrate was digested. Other evidence was cited earlier in this chapter. The digestion of ApApApA by L. acidophilus PDE follows a kinetic course which suggests strong binding of substrate to enzyme since only small quantities of ApApA and ApA are observed throughout the course of the reaction.[60] The L. casei PDE has been found to degrade first to oligomers of length 12 - 24, and then all at one to monomers.[108] This indicates that this enzyme may be tightly binding the substrate. It is interesting that this enzyme may have different modes of attack depending on the chain length of the substrate.

The kinetics of the above scheme can be represented by a set of differential equations, for (a) the free enzyme,

$$\dot{E} = -k_1 (E) (S_N) + k_3(ES_1) \tag{44a}$$

(b) the undegraded polymer

$$\dot{S}_N = -k_1(E)(S_N) \tag{44b}$$

(c) the enzyme substrate complexes

$$\begin{aligned} \dot{ES}_N &= k_1 (E)(S_N) - k_2(ES_N) \\ \dot{ES}_{N-n} &= k_2((ES_{N-n+1}) - (ES_{N-n})) \\ \dot{ES}_1 &= k_2(ES_2) - k_3(ES_1) \end{aligned} \tag{44c}$$

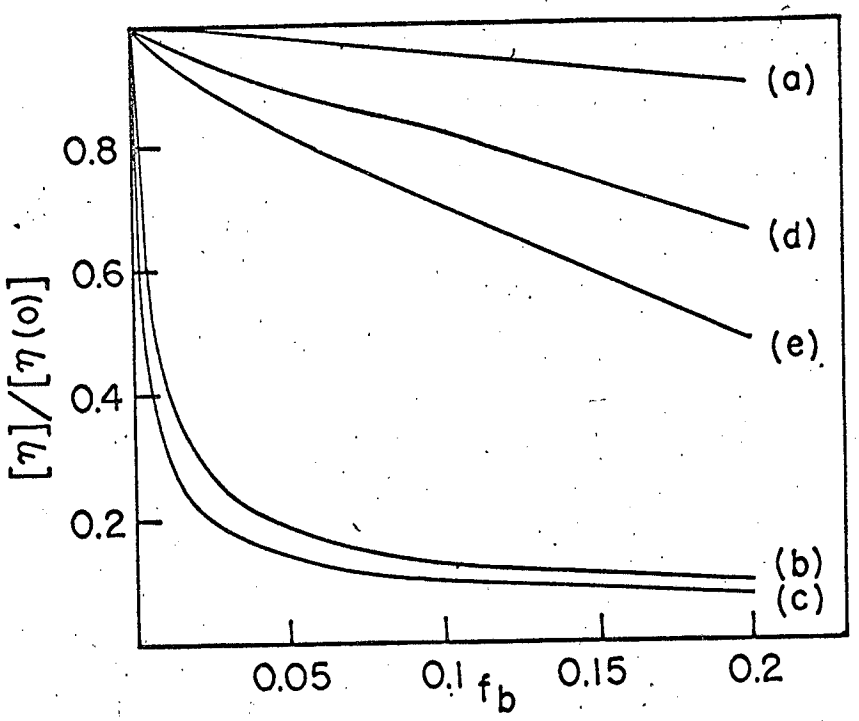


Figure 9. Change in viscosity of an RNA of chain length 1000, upon enzyme degradation.

a) Pure exoenzyme, calculated from equation (38).

b) Pure endoenzyme, calculated from equation (43a).

c) Pure endoenzyme, calculated from equation (43b)

d) Exoenzyme with 1% endo impurity, calculated from equations (38) and (43a).

e) Exoenzyme with 1% endo impurity, calculated from equations (38) and (43b).

and for (d) released monomer.

$$\dot{B}_n = k_2(ES_{N-n+1}) \qquad (44d)$$

$$B_N = k_3(ES_1)$$

Again we have made some implicit assumptions about the mechanism of the last step of the reaction. It is conceivable that B_N might be released simultaneously with B_{N-1} ($k_3 = \infty$) or even before. If it is desired to know the intermediate concentration of substrate S_2 through S_{N-1}, even though they are not free in the solution, one would stop the reaction by precipitating the enzyme and destroying all the complexes. The concentration of substrate would then be given by

$$"S"_{N-n} = ES_{N-n} \quad \text{where} \quad 2 \leq n \leq N-1 \qquad (45)$$

This approximation is necessitated by our assumption of only a single enzyme substrate complex for each step of the reaction.

The above series of differential equations contains one non linear equation. This must be linearized before solutions can be obtained. Two approximations can be made, and we will see that they lead to solutions with very different properties. In the limit of high enzyme concentration, we assume (E) is constant. This holds only if $E(0) >> S_N(0)$. We shall also set $ES_N(0) = 0$, since as we have shown previously the choice of this boundary condition has only a small effect on the resulting kinetics. If we let $k_1E = k_E$ and take the Laplace transform of the above set of differential equations (44b) through (44d), we have:

$$s\,\bar{S}_N - S_N(0) = -k_E\bar{S}_N \qquad (46a)$$

$$s\,\overline{ES_N} = k_E\bar{S}_N - k_2\,\overline{ES_N} \qquad (46b)$$

$$s\,\overline{ES}_{N=n+1} = k_2(\overline{ES}_{N-n+2} - \overline{ES}_{N-n+1}) \qquad n=2 \text{ to } N-1 \tag{46c}$$

$$s\,\overline{ES}_1 = k_2\,\overline{ES}_2 - k_3\,\overline{ES}_1 \tag{46d}$$

$$s\,\overline{B}_n = k_2\,\overline{ES}_{N-n+1} \qquad n=1 \text{ to } N-1 \tag{46e}$$

$$s\,\overline{B}_N = k_3\,\overline{ES}_1 \tag{46f}$$

The resulting set of linear equations can be easily solved. We find

$$\overline{B}_n = \frac{k_E k_2^{\,n}\, S_N(0)}{s(s + k_E)(s + k_2)^n} \qquad n=1 \text{ to } N-1 \tag{47}$$

$$\overline{B}_1 = \frac{k_E k_2^{\,n}\, S_N(0)}{s(s + k_E)(s + k_2)^{N-1}(s + k_3)} \tag{48}$$

In principle, one could expand the denominators of both transforms term by term. But for the moment we are interested in an estimate of the kinetic behavior of this model, so we will make a series of approximations which will give us results in much simpler form. We let

$$k_E = k_2 = k_3 = \gamma .$$

This will result in the solutions having only one exponential, $e^{-\gamma t}$, and does not really change the quantitative properties of the results, since k_E and k_3 appear only once while there may be hundreds of terms in k_2. With this simplification, we have

$$\overline{B}_n = \frac{\gamma^{n+1}\, S_N(0)}{s(s + \gamma)^{n+1}} \qquad n=1 \text{ to } N \tag{49}$$

Using the previously obtained result that

$$\mathcal{L}^{-1}\ \frac{1}{s(s+\beta)^n} = \frac{1}{\beta^n}\left(1 - \sum_{i=0}^{n-1} \frac{(\beta t)^i\, e^{-\beta t}}{i!}\right)$$

we have

$$B_n(t) = S_N(0)\left(1 - \sum_{i=0}^{n} \frac{(\gamma t)^i\, e^{-\gamma t}}{i!}\right) \tag{50}$$

It can easily be shown using the original differential equations that

$$ES_{N-n+1}(t) = S_N(0)\ \frac{(\gamma t)^n\, e^{-\gamma t}}{n!} = \text{"S"}_{N-n+1} \tag{51}$$

These equations have the same form as the results found previously for the random model of enzyme attack. If we compare $B_n(t)$ with the steady state random values, we find

$$B_n(\gamma t) = B_{n+1}(\beta t) \qquad \text{steady state random} \tag{52}$$

This difference is easily explained. The steady state model is equivalent to starting the reaction at earlier times, since the time needed for the concentration of ES_N to build up its steady state value is ignored. By contrast, we have

$$B_{2n-1}(\alpha t) = B_n(\gamma t) \quad \text{for the optimal random case.} \tag{53}$$

Thus, in the limit of high enzyme to substrate ratios, the kinetics of the tight binding enzyme model are very similar to the previously discussed random model. One further quantity of interest is the total rate of evolution of monomers.

$$\frac{dB_n(t)}{dt} = S_N(0)\gamma \frac{(\gamma t)^n e^{-\gamma t}}{n!} \tag{54}$$

For an infinite polymer we have

$$\sum_{n=1}^{\infty} \frac{dB_n(t)}{dt} = \sum_{n=1}^{\infty} S_N(0)\gamma \frac{(\gamma t)^n e^{-\gamma t}}{n!}$$

$$= S_N(0)\gamma (1 - e^{-\gamma t}) \tag{55}$$

Thus, γ can easily be determined from experiments.

13. Tight Binding Model--Low Enzyme Approximation

The second approximation we can use to solve the differential equations for the strong binding model is the assumption that S_N is a constnat. This will be approximately true if $S_N(0) >> E(0)$. In fact, we know that $S_N(t)$ must decrease with time slowly, and this can always be taken into account later. Thus, we let $k_1 S_N = k_s$. Taking the Laplace transform of equation (44), we have

$$s\,\overline{E} - E(0) = -k_s\overline{E} - k_3\overline{ES} \tag{56}$$

The other equations needed to solve the kinetics in this approximation are the same equations (46b) to (46f) of the constant enzyme approximation. The set of linear equations is solved as before, but, since this time it is not so straightforward, we shall give more details. Rewriting equation (56), we have

$$\overline{E} = \frac{E(0) + k_3\overline{ES}_1}{s + k_s} \tag{57}$$

Next, we combine equations (46b), (46c), and (46d) to give

$$\overline{ES}_{N-n+1} = \frac{k_s\, k_2^{n-1}\, \overline{E}}{(s + k_2)^n} \tag{58}$$

$$\overline{ES}_1 = \frac{k_2^{N-1} k_s\, \overline{E}}{(s + k_3)(s + k_2)^{N-1}} \tag{59}$$

Using equation (57), we can rewrite both of these as

$$\overline{ES}_{N-n+1} = \frac{k_s k_2^{n-1}}{(s + k_2)^n\,(s + k_s)} \left[E(0) + k_3 ES_1 \right] \tag{60}$$

$$\overline{ES}_1 = \frac{k_2^{N-1} k_s}{(s + k_3)(s + k_2)^{N-1}} \left[\frac{E(0) + k_3 \overline{ES}_1}{(s + k_s)} \right] \tag{61}$$

Now we again make the simplification $k_3 = k_2 = k_s = \delta$. As we mentioned previously, this approximation will only slightly change the results, while permitting us to write everything in much simpler form. We now solve equation (61) for $\overline{ES}_1$ and then substitute this value into equation (60).

$$\overline{ES}_1 = \frac{E(0)}{\delta} \left(\frac{\delta^{N+1}}{(s+\delta)^{N+1} - s^{N+1}} \right) \tag{62}$$

$$\overline{ES}_{N-n+1} = \frac{\delta^n E(0)}{(s+\delta)^{n+1}} + \frac{\delta^{n+1}}{(s+\delta)^{n+1}}\, \overline{ES}_1$$

$$= \frac{\delta^n E(0)}{(s+\delta)^{n+1}} + \frac{\delta^n E(0)}{(s+\delta)^{n+1}} \left(\frac{\delta^{N+1}}{(s+\delta)^{N-1} - \delta^{N+1}} \right) \tag{63}$$

To place this last equation in a simple form, we note that

$$\frac{\delta^{N+1}}{(s+\delta)^{N+1} - \delta^{N+1}} = \left(\frac{1}{\left(\frac{s+\delta}{\delta}\right)^{N+1} - 1} \right) \tag{64}$$

and use the expansion

$$\frac{1}{x-1} = \sum_{i=1}^{\infty} \frac{1}{x^i}$$

The result is

$$\overline{ES}_{N-n+1} = \frac{\delta^n E(0)}{(s+\delta)^{n+1}} \left[\sum_{i=0}^{\infty} \left(\frac{\delta^{N+1}}{(s+\delta)^{N+1}} \right)^i \right] \tag{65}$$

Thus we can write that

$$\overline{B}_n = \frac{\delta^{n+1} E(0)}{(s+\delta)^{n+1}} \left[\sum_{i=0}^{\infty} \left(\frac{\delta^{N+1}}{(s+\delta)^{N+1}} \right)^i \right] \tag{66}$$

and taking the inverse transform term by term we find

$$B_n(t) = E(0) \sum_{i=0}^{\infty} \left[1 - \sum_{j=0}^{n+i(N+1)} \frac{(\delta t)^j}{j!} e^{-\delta t} \right] \tag{67}$$

Thus far we have not taken into account the fact that there is only a finite amount of substrate. In order to do so we must truncate the infinite series. We shall take only $S_N(0)/E_N(0)$ terms. If this is non-integral it should be approximated by the nearest integral number of

terms plus a remainder which we shall ignore. We then note that each term is identical in form to the kinetics of evolution of some base in a much larger polymer according to the constant enzyme model derived previously. Thus we can rewrite equation (67) in the particularly simple form

$$B_n(t) = \frac{E(0)}{S_N(0)} \left[B_n(t) + B_{n+N+1}(t) + B_{n+2N+2}(t) + \cdots \right] \qquad (68)$$

where there are $S_N(0)/E(0)$ terms in the sum and we replace γ by δ. It is easily shown that this equation has the proper limits at $t=0$ and $t=\infty$. Thus in the limit of low enzyme to substrate ratios the kinetics of degradation of an oligomer is the same as the degradation of a much larger polymer containing the repeating oligomer sequence (spaced by one imaginary residue).

As before, we want to find an expression for the total rate of monomer evolution from an infinite polymer.

$$\sum_{n=1}^{\infty} \frac{dB_n(t)}{dt} = \frac{E(0)}{S_N(0)} \sum_{n=1}^{\infty} \sum_{i=0}^{S_N(0)/E(0)} \frac{dB_{n+i(N+1)}(t)}{dt} \qquad (69)$$

Since each $(dB_j)/dt$ is simply $S_N(0)\delta(\delta t)^j e^{-\delta t}/j!$ we have

$$\sum_{n=1}^{\infty} \frac{dB_n(t)}{dt} = E(0)\delta e^{-\delta t} \sum_{n=1}^{\infty} \sum_{i=0}^{S_N(0)/E(0)} \frac{(\delta t)^{n+i(N+1)}}{(n+i(N+1))!} \qquad (70)$$

And if N is large and t small, we can neglect all but the first term of the second sum. We find that

$$\sum_{n=1}^{\infty} \frac{dB_n(t)}{dt} \cong E(0)\delta(1 - e^{-\delta t}) \qquad (71)$$

A determination of the total rate of monomer evolution will thus permit evaluation of δ. As before, we can also write down an expression for $"S"_n(t)$ which is the concentration of substrate we would find if we stopped the reaction and dissociated all the complexes.

$$"S"_n(t) = E(0)/S_N(0) \left["S"_n(t) + "S"_{n+N+1}(t) + "S"_{n+2(N+1)}(t) \cdots \right] \qquad (72)$$

where each of the $S_N(0)/E(0)$ terms has the same form as the high enzyme to substrate case, but with γ replaced by δ. This shows that here the distribution of chain lengths is no longer a simple Poissonian, but is much broader.

If we had an exoenzyme which we knew worked by the tight binding mechnaism, we would probably find it difficult to use it to get much sequence information. The best we could do would be to use the enzyme in very great molar excess over the substrate. If we could analyze the reaction mixture all of the same conclusions would apply which were derived for the steady state case of the random model. If we accidently used the enzyme in too low a concentration we would get what would appear to be an almost random distribution of monomers. Thus if an exoenzyme digestion of an RNA gives a random evolution of monomers it is necessary to check the homogeneity of the substrate and also the nature of the enzyme.

14. How to Decide Which Mode of Attack is Occurring

Since exoenzymes which work via a random mode of attack are potentially better for sequence studies than those which are tightly bound to the substrate, it is desirable to have a sensitive test of the mode of enzyme action. There are several simple ways in which a random enzyme could be distinguished from a strong binding one. These methods all arise from the fact that the kinetics of degradation by the two modes of attack are very different in the limit of low enzyme concentration.

If we study the degradation of any oligonucleotide substrate with a great excess of enzyme, the various monomers will be given off in an orderly fashion and unless the kinetics are examined in great detail no accurate estimate of the mode of enzyme attack can be made. But in the low enzyme limit, the tight binding enzyme will evolve a given nucleotide in an oscillatory fashion. A simple example is the degradation of an oligomer of the type $(Ap)_{10}G$ by venom PDE (which degrades stepwise from the right). Such a substrate could be prepared in principle in the following manner. RNA is hydrolyzed to completion with T1 ribonuclease, and the resulting mixture of oligonucleotides is separated by chain length. The 11-mer fraction is then hydrolyzed to completion with pancreatic RNAase. The resulting mixture is again separated by chain length. The only oligomer with chain length 11 remaining is $(Ap)_{10}G$. If there is three times as much substrate as enzyme, and the enzyme works by the tight binding mode, the kinetics of appearance of pG will be given by

$$[pG] = 1/3\ [B_2(\gamma t) + B_{14}(\gamma t) + B_{26}(\gamma t)] \tag{73a}$$

where the $B(\gamma t)$'s are taken from the values calculated from the steady

state case of the random model, or

$$[pG] = 1/3\ [B_1(\gamma t) + B_{13}(\gamma t) + B_{25}(\gamma t)] \qquad (73b)$$

where the $B(\gamma t)$'s are taken from the values of the high enzyme limit of the tight binding model. These two expressions for [pG] are identical.

In contrast, if the enzyme acts by the random mode of attack, then [pG] is given either by

$$[pG] = B_1(\beta t) \qquad (74a)$$

where $B_1(\beta t)$ is the steady state case of the random model, or by

$$[pG] = B_1(\alpha t). \qquad (74b)$$

Here B_1 is calculated from the optimal case. The correct value of [pG] will fall somewhere inbetween these two expressions.

In Figure 10 we have compared the values of [pG] calculated by equations (73b) and (74b). It is easily seen that the kinetics of appearance of pG are entirely different for the two models, and thus they could be easily distinguished by such an experiment. A simple check would be provided by changing the substrate to enzyme ratio to, say, 6:1 and performing another hydrolysis. This should have little effect other than a change in overall rate if the enzyme acts by the random mode, but a pronounced change will be observed if the enzyme acts in a tight binding fashion.

An alternative method of distinguishing between the two model of attack is based on the fact that the distribution of substrates of various chain lengths is very different for the two models. Intuitively, one can see that the concentration of substrates of intermediate chain length will

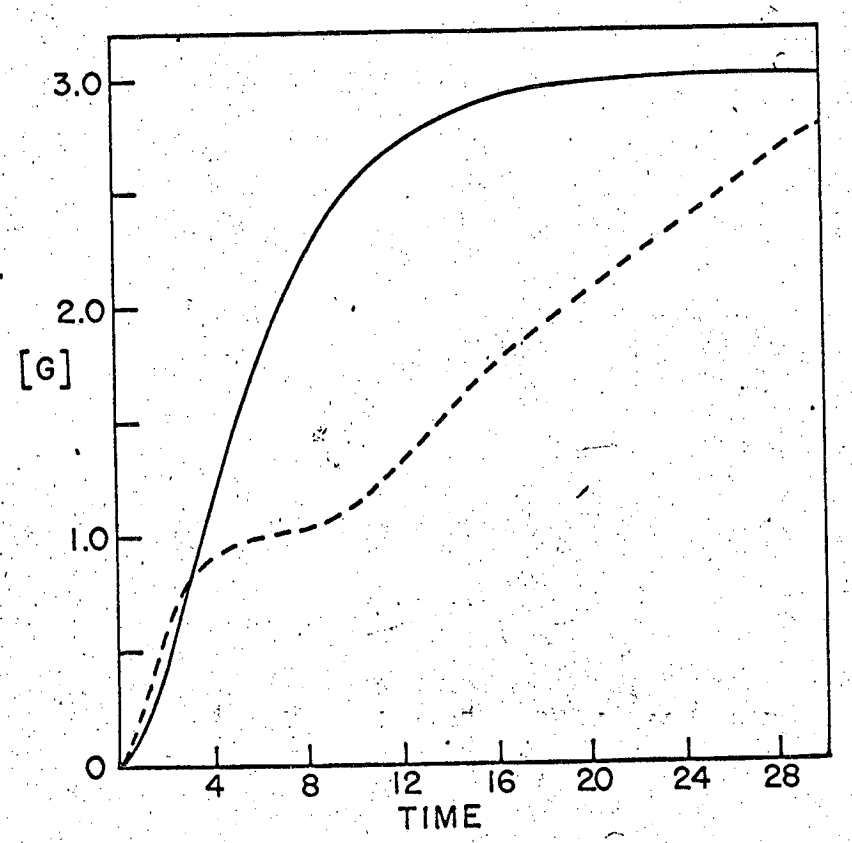


Figure 10. Comparison of the mode of action of exonucleases acting by random and tight binding mode. $S_N(0) = 3E(0)$. Evolution of pG from $(Ap)_{10}G$ by a 5' phosphodiesterase.

- - - - - - tight binding enzyme.

———— random enzyme.

be much lower in the case of strong binding.

For the steady state case of the random enzyme model, the distribution of substrates as a function of time can easily be calculated. Recall that for this model

$$S_{N-n+1}(t) = S_N(0) \frac{(\beta t)^{n-1} e^{-\beta t}}{(n-1)!} \tag{75}$$

The maximum value S_{N-n+1} can obtain is found by setting its time derivative equal to zero.

$$\frac{dS_{N-n+1}(t)}{dt} = 0 \Rightarrow \beta t_m = n-1 \tag{76}$$

Thus we find that the maximum value of S_{N-n+1} relative to the starting concentration of substrate is

$$\frac{S_{N-n+1}(n-1)}{S_N(0)} = \frac{(n-1)^{n-1} e^{-(n-1)}}{(n-1)!} \tag{77}$$

For large n we can use the well known approximation that

$$n! = \sqrt{2\pi n}\,(n/e)^n.$$

Substituting this into the above equation we have

$$\frac{S_{N-n+1}(n-1)}{S_N(0)} \simeq \frac{1}{\sqrt{2\pi\,(n-1)}} \quad \text{for large } n. \tag{78}$$

For small n we have tabulated the following values.

n	1	2	3	4	5
$\frac{S_{N-n+1}(n-1)}{S_N(0)}$	1	.368	.270	.225	.192

In terms of the maximum value of S_{N-n+1} it can easily be shown that the shape of the distribution at $\beta t = n-1$ is, for the steady state random case,

$$\frac{S_{N-(n+i)+1}}{S_{N-n+1}} = (n-1)^i \frac{(n-1)!}{(n-1+i)!} \qquad (79)$$

Now in the case of tight binding, the expression for $S_{N-n+1}(t)$ becomes more complex. In the limit of low enzyme concentration, we use equation (72). This function will have $S_N(0)/E(0)$ relative maxima, at $t = n+i(N+1)$ where $i = 1$ to $S_N(0)/E(0)$. Thus, as we expected, the maximum concentration of intermediate length substrates is much smaller in the tight binding case.

To be able to compare the distribution functions of $S_{N-n+1}(t)$ for the two models more directly we have calculated the distribution of fragments from a hypothetical 20-mer. These results are shown in Figure 11. They are computed for the time when half of the monomers have been released. For the random model the steady state formulation was employed, so this time is $19/\beta$. For the tight binding case, we assumed a substrate to enzyme ratio of 5:1. Thus the distribution is approximately equal to the distribution of fragments from a 100-mer in the steady state random model when $t = 49/\beta$, multiplied by 1/5. From Figure 11 it is apparent that the concentration of partially degraded chains in the tight binding case is much less than the random case, and the distribution is also much broader. In the random case the concentration of undegraded polymer has fallen to zero, while in the tight binding case more than 40% of the original polymer is still whole. The difference between these two distributions could easily be detected by a chromatography which separates oligomers according to chain length.

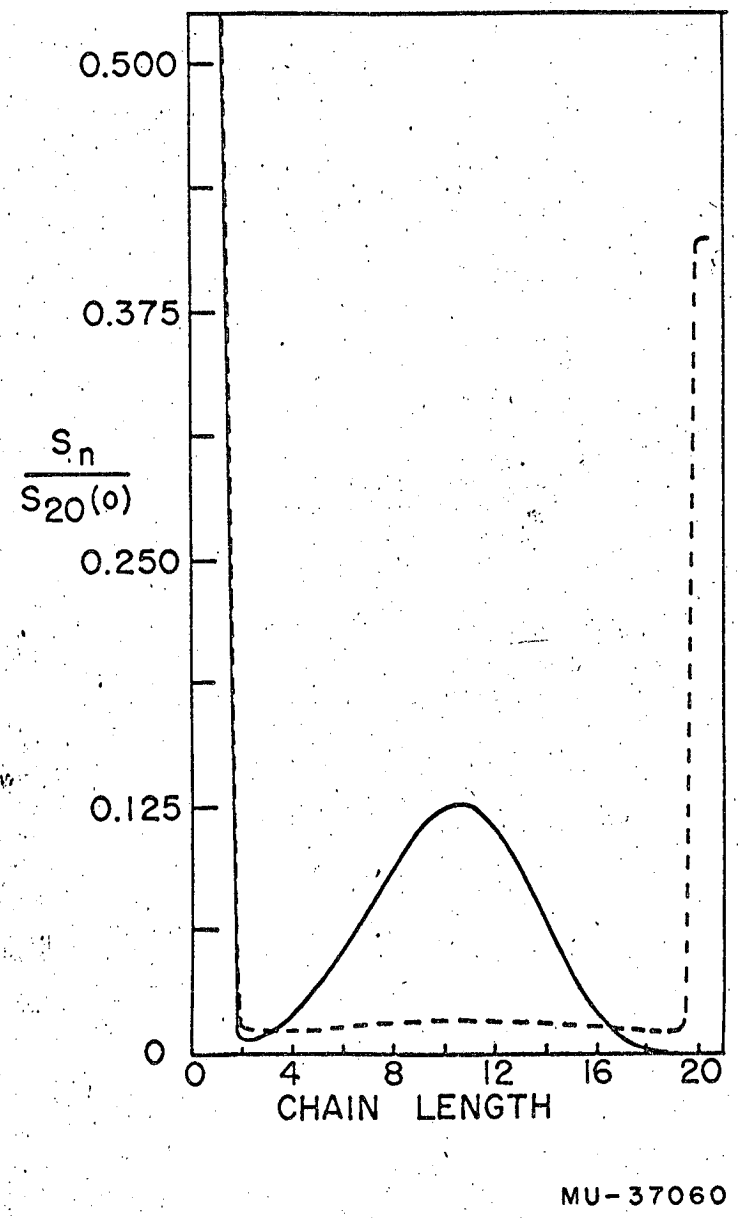


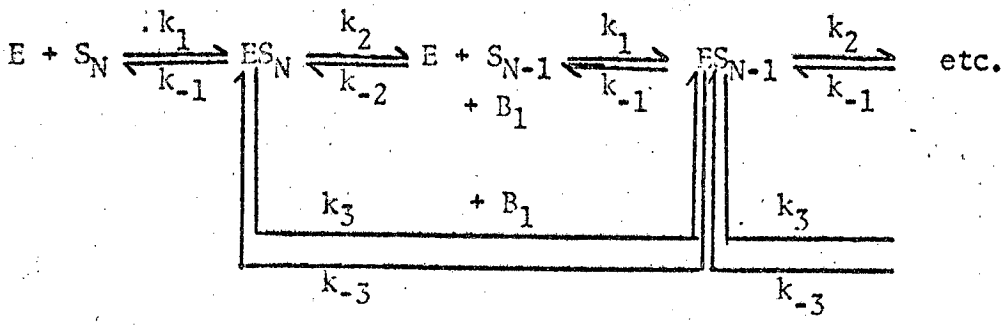
Figure 11. Distribution of oligomer fragments from the exonuclease hydrolysis of an oligonucleotide containing 20 monomer residues when half of the phosphate linkages have been broken. $S_N(0) = 5E(0)$.

- - - - - - tight binding enzyme.

———— random enzyme.

15. Future Work

In actual practice, the kinetics of exoenzymes will probably fall somewhere between the two models we have discussed. It can be represented by the following reaction scheme where the relative values of k_1, k_{-1}, k_2, k_{-2}, k_3, and k_{-3} probably depend on chain length.



Even if the approximation is made that k_{-2} and k_{-3} are zero, the system of differential equations which results from this scheme is too complicated to solve conveniently, and the results would require too many parameters to fit to the existing experimental data. More accurate experimental data are needed to decide whether or not the general scheme must be used, or instead, if conditions can be found where either the tight binding or random mode of attack predominates.

It is interesting to speculate on the types of mechanism of action which can lead to the tight binding case. It may be that the enzyme has two sites. One of them binds the substrate at some nucleotide other than the terminal residue, while the actual cleavage is performed at a second site. Then the substrate is advanced as if on an assembly line. This would explain why some phosphodiesterases, like RNAase II from *E. coli*, cannot degrade oligomers shorter than 5 or 6 residues.[210] If

in fact the biological role of phosphodiesterases is to degrade mRNA as it comes off the ribosome, as has been suggested by several authors (see Section 2 of this chapter), this mode of action is quite understandable. Another possibility of great interest from the point of view of kinetics is that the tight binding mode of enzyme action may be a manifestation of a diffusion controlled reaction. The enzyme may actually release the polymer as in the random mode of attack, but, because both reactants diffuse very slowly (and might even become entangled), the probability of a collision between the two species again is very high. Thus the strong binding mode kinetics may simply be counting the number of times a substrate is hit before it is able to diffuse away from the enzyme.

There are several ways in which the two mechanisms may be distinguished. If there really is tight binding of the substrate to the enzyme, the details of kinetics should only depend on $S_N(0)/E(0)$, and would be relatively independent of the actual values of the two. The total rate of course would be linear in whichever is smaller. But if the apparent tight binding is due to diffusion, then the random mode of attack should become favored as the concentrations of $S_N(0)$ and $E(0)$ are increased. This would only affect the details of the kinetics if diffusion controlled reactions were playing an important role.

There is another, much more complicated type of experiment which could detect even a small amount of dissociation of the enzyme substrate complex. The substrate is a short oligonucleotide of the sequence $Ap(Bp)_nC$. A large excess of this substrate is exposed to a mixture of two enzymes, a 5' PDE and a 3' PDE, both of which are thought to behave according to the strong binding mode. If both enzymes bind their sub-

strates irreversibly, then the only products of this reaction after total hydrolysis are A, Ap, pB, Bp, pC, and C. But if any partially degraded substrate manages to escape from its enzyme some nucleoside, B, will appear as a product. In order for this experiment to be a success, both enzymes must have only minute phosphatase activities. A variation of this experiment would be to bind a 5' PDE covalently to a 3' PDE. A study of the reaction products from the hydrolysis of $Ap(Bp)_nC$ by such a composite would enable the detection of even a slight tendency of the enzyme substrate complex to dissociate. A careful examination of the kinetics might also enable the determination of the number of times a substrate changes enzymes. Any study of this type, as well as the elucidation of the structure and properties of these enzymes, is contingent upon the enzymologists' ability to prepare pure and homogeneous samples.

The role that exoenzymes play in future sequence studies of RNA's remains to be seen. An entirely different method of using exoenzymes to obtain sequence information has recently been used by Holley and his co-workers.[99] This method will probably receive much attention in the future. The oligomer is partially digested with a 5' PDE, and the resulting set of oligomers is separated according to chain length by chromatography. Then each fragment is hydrolyzed with alkali and either the terminal residue, which appears as a nucleoside, is determined or the base composition of the whole fragment is analyzed. This provides a simple and convenient way of reconstructing the sequence of the oligomer providing that a chromatographic separation of the fragments can be carried out. Thus, this method is limited at present to sequence determination of oligonucleotides. The determination of the terminal

sequence of a large polymer by this method would not be feasible. The method used by Holley is limited only by the width of the distribution of oligomers formed, but as we have shown the distribution does not broaden that rapidly. Thus the sequence of a 20-mer could be determined in this manner if enough material was used so that the terminal residue of the 10-mer formed after half of the substrate has been digested could be detected. Since the maximum fraction of 10-mer is about .13, only 1% of the bases of substrate can be isolated as the 10th residue from the 3' end. Part of the data from an actual sequence determination, taken from Holley, Madison, and Zamir[99] is shown in Figure 12. The oligomer was GpGpGpApGpApGpU. An interesting sidelight of this experiment is that the concentrations of GpGpGpApG and GpGpGpApGpApG are much larger than would be expected from random exo attack. This is an indication that the ApG bond is cleaved much more slowly than GpA. This is in contrast to the results reported by most authors, which were discussed in Section 2 of this chapter, that the kinetics of exoenzyme attack do not depend on the particular nucleotide being cleaved. We hope that further experiments will clarify this discrepancy.

If exonucleases can be found in the future which do indeed show total base specificity, the methods we have discussed will increase their effectiveness substantially. Consider, for example, an exoenzyme which degrades a polymer chain until it reached a C and then stops, unable to break an NpC bond. It may be possible to produce this kind of specificity by chemically modifying some of the nucleotides. Then the C is removed by terminal chemical degradation and the hydrolysis is continued. Thus a simple analysis of the total number of monomers released in each pulse will permit the location of all the C's to be mapped out.

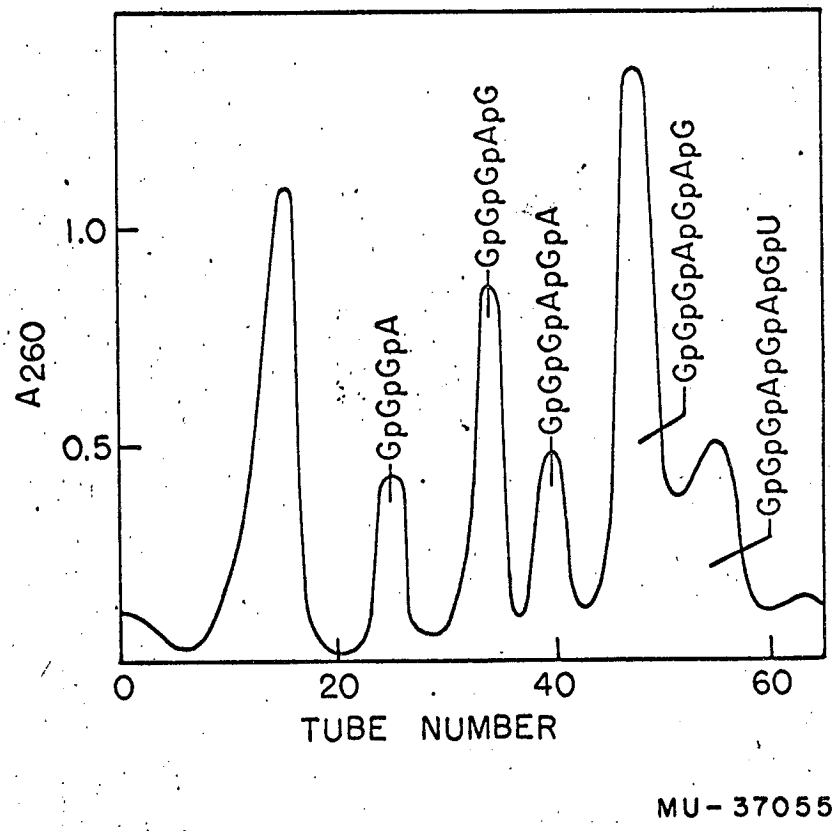


Figure 12. Distribution of oligomers from the partial hydrolysis of GpGpGpApGpApGpU by venom phosphodiesterase. Adapted from Holley, Madison, and Zamir.[99]

Given three other enzymes of similar inclination but different specificities; given an automatic method of analysis (in this case a simple pH-stat); given a solid state terminal degradation, then the problem of sequence determination is solved. Until such good fortune comes our way we shall have to be content with today's rather unspecific exonucleases.

In a more practical vein, the conformational dependence of some of the exonucleases could be profitably used to simplify the determination of sequences. If one believes that a given RNA has only one unique conformation (at any temperature, pH and ionic strength), then an exoenzyme should degrade a single stranded end until it reaches the first base paired residue. Then all activity would halt until the experimenter changes either temperature, pH, or ionic strength. Then the enzyme will take off again until it reaches the next blockade. By such a procedure the distribution of substrates could be kept much narrower and the corresponding evolution of monomers much sharper. Thus, while all our previous discussion has dealt with non-specific exoenzymes, the experimenter seriously interested in determining the sequence of a nucleic acid would do well to use the most specific enzyme he can find. But he should beware since a partially specific exonuclease is worse than no specificity at all.

III. OPTICAL PROPERTIES OF OLIGONUCLEOTIDES

1. Oligomer Models for RNA

To understand the conformation of RNA in solution is a major objective for the physical chemist interested in biological molecules. It will provide a start in elucidating the conformation of the various kinds of RNA *in vivo*. A knowledge of the three-dimensional geometry of RNA molecules will be useful in correlating their biological properties with their chemical structure. While physical chemical studies will not provide all the answers, they will greatly assist the concurrent work of the biologist and biochemist.

A detailed description of the mechanism of protein synthesis is almost certainly contingent upon our understanding the conformation of the three types of RNA involved. A study of the conformation of rRNA is a prerequisite to understanding how mRNA and sRNA bind to the ribosome. How each amino acyl RNA synthetase recognizes the proper sRNA may very well depend on the conformation of the RNA, since small differences in sequence could be amplified into large differences in conformation.[255] Differences in conformation may be responsible for the fact that some mRNA's are long lived while others have a half life of only two minutes *in vivo*.[255] Since it is suspected that nucleases are responsible for the breakdown of RNA *in vivo*, the differential susceptibility of various RNA conformations to nuclease attack may play a very important biological role. It is possible that the conformation of RNA attached to histones helps to determine the mode of action of these regulatory proteins. Therefore the transcription and translation of the genetic

message may be intimately connected to the conformation of the macromolecules involved. Thus there is a strong incentive to determine the conformation of RNA.

Because of the complexity of any native RNA molecules, it seemed to us that it would be helpful to first study small oligonucleotides. We could then use the properties of these compounds to predict the conformation of larger RNA's. The simplest possible approach is to try to understand the properties of polymers in terms of monomer properties. If we could hope to be able to do this, then we should be able to predict the properties of dinucleoside phosphates or dinucleotides from the four monomers. This would not guarantee that monomer properties can explain polymers, but it is a necessary condition. Unfortunately, attempts to predict the optical properties of dimers from monomers have thus far been mostly very disappointing.[25] The optical properties of polymers like poly A or RNA bear little resemblance to the monomers. And there are, as yet, no practical ways of predicting the hydrodynamic properties of a large polyelectrolyte with a conformation that is highly irregular.

At the next level of complexity is an attempt to predict the properties of RNA in terms of the properties of the 16 dinucleoside phosphates. Here the trouble arises that there are so many different observations needed before the polymer can be predicted. The polymer properties would be an average over all the dimer properties, and agreement between experimental and predicted data might be due to fortuitous canceling of different dimer properties. Thus it would be advisable to test our ability to use dimer properties to predict the properties of, say, trinucleoside diphosphates. Here fewer pieces of data are needed as input and the result will be a much more sensitive test of the corre-

lation (if any) between the conformations of dimers and trimers. If we can predict any trimer properties from dimer properties it will mean that these observables are not very sensitive to next-nearest neighbor interactions. If the trimers can be understood in terms of their component dimers, the next step will be to try to predict the properties of polymers.

Thus we decided to study the properties of some trinucleoside diphosphates as the simplest models on which to test our understanding of dimers. In view of the triplet code, the properties of trinucleoside diphosphates might be especially interesting since one would be studying a word of a genetic message. There is some evidence that the conformation of trinucleotide fragments of RNA may play an important role in the translation of the code on the ribosome.[35] An important factor in our decision to work with trinucleoside diphosphates is that as the chain length of an oligonucleotide increases arithmetically the difficulty of preparing it in useful quantities increases geometrically.

There are many physical measurements capable of providing information about the conformation of a molecule in solution. These include UV spectroscopy, optical rotatory dispersion (ORD), circular dichroism (CD), magnetic resonance, light scattering, viscometry, sedimentation, small angle X-ray scattering, IR spectroscopy, hydrogen exchange, and titrimetry. But most of these techniques require a relatively large amount of sample. This immediately limits the techniques which should be chosen for a preliminary study of the conformation of oligonucleotides which can only be obtained in small amounts. Furthermore, many of the techniques are limited by the fact that some oligonucleotides

are only sparingly soluble in water. Fortunately, the optical properties of nucleotide containing compounds are very intense. Typically, an oligonucleotide has an extinction coefficient at 260 mμ of more than 10^4 per nucleotide residue, and a molar rotation at the longest wavelength trough of 10^4 degrees per residue. Thus a study using either UV spectroscopy or ORD can be carried out using less than milligram quantities of material.

There was already much evidence that the ORD and related optical properties of poly A,[19,95] poly C,[54] sRNA,[116] and ApA[147,253,246] are strongly dependent on the conformation. We had every reason to believe that the optical properties of trimers would also be strongly dependent on conformation. Furthermore, an extensive study of these optical properties was already being carried out in our laboratory. It was expected that all of the 16 dinucleoside phosphates and the four mononucleotides and mononucleosides would have been studied by the time our work on the trinucleoside diphosphates was completed. Thus we would be in a position to critically test our ability to correlate dimer and trimer properties. All of the above factors were instrumental in our decision to restrict our study of the conformation of trinucleoside diphosphates to their UV optical properties.

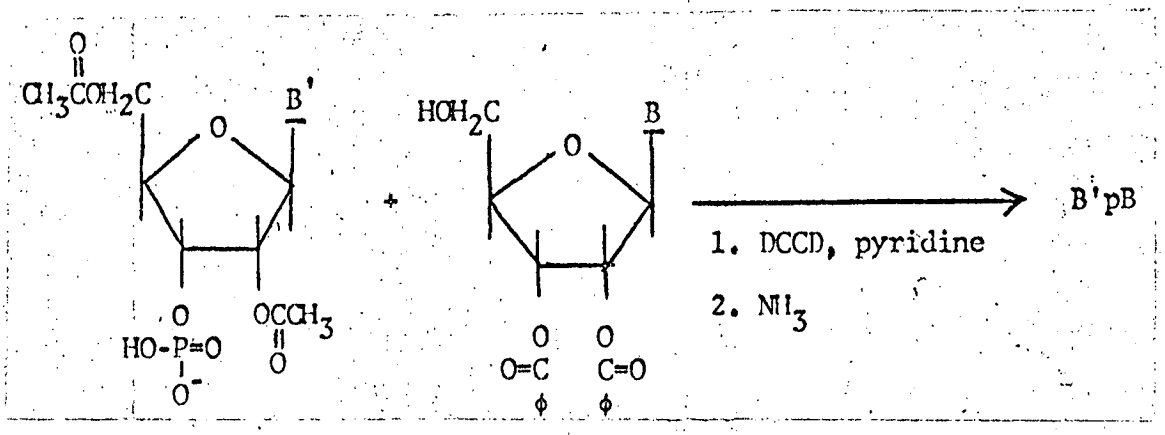
Our second goal is to determine whether the nucleotide sequence of an oligonucleotide can be determined by using only measurements of the optical properties. All of the methods of base sequence determination discussed in Chapter II degrade the molecule being sequenced. A more desirable approach would leave the oligonucleotide intact. Since ORD is known to depend strongly on interaction between neighboring chromaphores,[25] it seemed that ORD was a likely physical method to yield sequence information.

2. Methods of Preparing Oligonucleotides

For the reasons outlined in the previous section, we decided to prepare several trinucleoside diphosphates. We wanted a set with varying base composition and at least one pair of sequence isomers. We also needed a method which would yield of the order of 1 mg of each trimer in reasonable purity. In this section we shall discuss the various ways in which trinucleoside diphosphates can be prepared, and we shall comment on the relative merits and disadvantages.

A natural approach to the preparation of any small biochemical compound is organic synthesis. The techniques of synthesizing oligonucleotides have been developed to a large degree by Khorana and his collaborators. They have had much success in preparing large amounts of many oligodeoxyribonucleotides, and in spite of the considerable difficulties involved, they have now developed methods which are capable of synthesizing any oligoribonucleotide. For brevity, we shall discuss only the most recent methods.

The general scheme for the synthesis of oligomers starts with the appropriately protected derivatives of monomers or oligomers which are coupled together using a condensing agent. Some protective groups are removed, and the process is then repeated. The major difficulty is that the reactions are not completely specific, and thus a large number of different products are formed which usually must be separated by ion exchange chromatography. The reactions currently used by Lohrmann and Khorana are shown on the following page,[136]



where B is any one of the following compounds: N,N'-dibenzoyladenine, N-benzoylcytosine, N-acetylguanine, or N-benzoyluracil. B' can be either N-acetylguanine or N-acetylcytosine. DCCD is dicyclohexylcarbodiimide. In this way, eight of the 16 possible dinucleoside phosphates were readily made in large amounts. To synthesize longer oligomers the 5'-O-acetyl derivative on B' is removed and the protected dimer is then ready to be coupled to another protected nucleotide. Lapidot and Khorana have prepared the tetranucleotide UAUU in this manner.[120] In a recent paper, Soll and Khorana have shown that a nucleoside 2',3'-cyclic phosphate can be used instead of the protected nucleoside.[214] Through the same reactions as above, this yields the dinucleotide B'pBp. CpUp, ApUp, and IpUp have been prepared in this way. The major advantage of any of the synthetic methods is that in principle any oligonucleotide can be made, and the often-times troublesome separation of very similar isomers is avoided. The major disadvantage with synthesis is that one prepares only one compound at a time. Once the appropriately protected derivatives are commercially available these difficulties can be overcome. For the above reason, we have thus far not strongly considered using any of the above methods.

In lieu of synthesis, the only other conceivable method of obtaining oligonucleotides seems to be degradation of natural or synthetic polynucleotides followed by separation of the fragments formed. We know of no natural source of small oligonucleotides themselves, although there seems to be no reason in principle why they could not exist in some organism in quantities large enough to merit consideration. The choice of a polynucleotide to be used for degradation to oligomers is limited by many factors. The first, and most important, is that any degradation is bound to produce a small yield of any oligomer, and thus the starting material should be available in large amounts. In order to limit the number of possible fragments, the RNA we use should have as few different nucleotides as possible. This immediately eliminates mixed sRNA unless it is desired to make oligomers containing all of the unusual bases contained in a typical preparation os sRNA. Ideally, one would want to use a polymer like synthetic poly AC, but the synthetic copolymers are difficult to prepare in very large amounts. In most studies, therefore, rRNA or viral RNA have been degraded in order to produce significant amounts of oligomers containing the normal bases, A, U, C, and G.

There are several ways in which RNA can be conveniently broken down into small oligonucleotides. The choice of a degradative method will have a profound influence on the complexity of the mixture of products formed. Hydrolysis with base (NaOH or $Ba(OH)_2$) has long been used to degrade RNA to its component mononucleotides. If the reaction is not allowed to go to completion, a mixture of oligomers will be formed. In principle, this will include every possible sequence and base composition of every chain length. Thus there will be four mononucleotides,

16 dinucleotides, 64 trinucleotides, etc. The yields will be low, and can be adjusted by controlling the extent of hydrolysis. If desired, the terminal 2' and 3' phosphates formed by this reaction can be removed using E. coli alkaline phosphatase. In this manner, Dimroth and Witzel were able to obtain all of the 16 dinucleotides.[42] If the RNA contains any 2'-O-methylnucleotides, it cannot be hydrolyzed to completion with base. Sequences of the type NxpNp, where x means 2'-O-methyl, are resistant to base hydrolysis presumably because the intermediate in the reaction, the 2',3'-cyclic nucleotide, cannot form.[124] In this way, quite a few alkali-resistant trinucleotides have been prepared.[117]

The method which has been most frequently used to degrade RNA to oligonucleotides is complete hydrolysis with pancreatic RNAase. This enzyme has the advantage that it can cleave RNA to yield chains ending in only 3'-cytidylic acid or 3'-uridylic acid. Another attraction of pancreatic RNAase is that it has a very high activity, so that the hydrolysis can be completed in a short time. This permits contamination from bacterial growth or from non-specific nucleases that are sometimes present in RNA preparations to be avoided. Complete hydrolysis of a long RNA with pancreatic RNAase will yield products of the following type:

Monomers	Up, Cp
Dimers	ApUp, ApCp, GpUp, GpCp
Trimers	ApApCp, ApApUp, ApGpCp, GpApCp, ApGpUp, GpApUp, GpGpCp, GpGpUp
etc.	

In general, there will be 2^n oligomers with chain length n. If the original RNA contained equal amounts of each of the four bases, and the se-

quence is random, the following yields can be expected:

1/8 of the starting RNA as each monomer,
1/16 of the starting RNA as each dimer,
3/128 of the starting RNA as each trimer,
5/16 of the RNA as higher oligomers.

Further details are given in Appendix 4. If the specificity of pancreatic RNAase is taken for granted, these hydrolyses can then be used to test for deviations from random sequence in the starting RNA.[58,8,264,220] Oligonucleotides from pancreatic RNAase hydrolysis of RNA have been isolated by many workers, including Staehelin,[218], Stanley,[222] and Rushizky and Sober.[190]

Recently, T1 RNAase has become available.[51] This enzyme can cleave RNA only after G, and must produce oligonucleotides which end in 3'-guanylic acid. A total hydrolysis of RNA with T1 RNAase will thus yield the following products:

Monomers	Gp
Dimers	ApGp, CpGp, UpGp
Trimers	ApApGp, ApCpGp, CpApGp, ApUpGp,
	UpApGp, UpUpGp, CpUpGp, UpCpGp, CpCpG
etc.	

In general, there will be 3^{n-1} oligomers with chain length n. If the original RNA was equal molar in the four common bases, and if the sequence is essentially random, complete digestion will result in the following yields:

1/16 of the starting RNA as Gp,

3/96 of the starting RNA as each dimer,

3/256 of the starting RNA as each trimer,

73.8% of the RNA as higher oligomers.

Thus the yield of individual monomers, dimers, or trimers is one half the yield of the corresponding pancreatic RNAase oligomer. Further details are given in Appendix 4. Most of the T1 dimers, trimers and tetramers have been isolated by Rushizky and Sober.[188,189] The C specific endonuclease from E. coli[1] will give similar yields of fragments. Lee, Ho, and Gilham have been able to make pancreatic RNAase specific for C by chemically modifying the RNA. By using RNA treated in this way, they have been able to isolate the nine trimers ending in Cp.[123] As other specific nucleases are found they will permit many more oligomers to be prepared in this way. But recent developments, discussed later, have made degradative methods practically obsolete.

The major difficulty with any of the degradative methods of oligomer preparation lies in the separation into individual components of the extremely complex mixture formed. However, as an off-shoot to the problem of sequence determination, many elegant and powerful methods of separating oligonucleotides have been developed. These have recently been reviewed by Staehelin.[221] With such a complex mixture of compounds as that obtained by degradation of an RNA, a systematic approach to the problem of separation is often very fruitful. Thus, methods have been developed which separate RNA fragments according to chain length, independent of the base composition or sequence. The most widely used is the ion exchange chromatography on DEAE cellulose in the presence of seven molar

urea developed by Tomlinson and Tener.[241] Recent modifications have substituted DEAE sephadex, which offers higher resolution in some cases.[185] The resulting mixture of oligomers of a given chain length can then be separated according to composition either by ion exchange chromatography,[190] or electrophoresis. An attractive alternative is the use of polynucleotide cellulose columns which potentially can separate oligomers differing only in sequence.[72] An alternative to two sequential separations is the use of a two dimensional fractionation resulting in an oligonucleotide map.[187,193] The mapping techniques currently in use have only limited applicability for large preparative separations. We have found, though, that the electrophoresis step commonly used in such separations can be scaled up by using the Spinco continuous electrophoresis apparatus. At the present time, however, the chromatographic methods seem to offer higher resolution.

Either of the two nucleases discussed in the previous paragraphs can be used to catalyze the synthesis of oligonucleotides. A large excess of the intermediates in the degradative reaction are mixed in the presence of the enzyme, and the reaction is allowed to occur. The enzyme is killed before the reaction mixture is diluted. The resulting equilibrium mixture of oligomers is separated by any of the methods previously described. Using T1 RNAase, the following reaction has been reported to occur in good yield.[51]

$$G{>}p + N \xrightarrow{\text{T1 RNAase}} GpN \qquad \text{where N is either A, C, U, or G}$$

In principle, trinucleoside diphosphates could be made by either of the following two reactions.

$$G{>}p + NpM \xrightarrow{\text{T1 RNAase}} GpNpM \qquad \text{where N is preferably not G,}$$

$$NpG>p + M \xrightarrow{\text{T1 RNAase}} NpGpM$$ where N is preferably not G.

If N is G the dinucleoside phosphate could be degraded, and a more complex mixture of products might result. A total of 12 different trimers could be made by each reaction scheme and, thus, 24 of the 64 possible trimers could be prepared in this way. The major difficulty lies in the fact that dinucleoside cyclic phosphates have thus far not been obtained in significant amounts. If pancreatic RNAase is used there are many more possibilities for synthesis. Nirenberg and his coworkers[164] have reported making many trinucleoside diphosphates using a derivative of pancreatic RNAase. Their synthesis involves either one of the following two reactions,

$$U>p + NpM \xrightarrow{\text{Panc. RNAase}} UpNpM$$

$$C>p + NpM \xrightarrow{\text{Panc. RNAase}} CpNpM$$

where N is preferably (but, apparently, not always necessarily) A or G. Sixteen possible trimers can be made in this way. In addition, if dinucleoside cyclic phosphates can be prepared, the following two reactions would be expected to yield trinucleoside diphosphates:

$$NpU>p + M \xrightarrow{\text{Panc. RNAase}} NpUpM$$

$$NpC>p + M \xrightarrow{\text{Panc. RNAase}} NpCpM$$

where N is preferably A or G. This leads to 16 more possible trinucleoside diphosphates and thus a total of 32 of the possible 64 can, in principle, be made by using pancreatic RNAase in a synthetic mode.

The methods discussed above would seem very attractive if it were not for a new similar method using the polynucleotide phosphorylase (PPase) obtained from Micrococcus lysodeikticus. It has long been known that this enzyme readily effects the synthesis of polynucleotides from nucleoside diphosphates according to the reaction[224]

$$n\ ppN \xrightarrow{PPase} (pN)_n + nP_i\ .$$

It was later shown that this synthetic property was, in fact, characteristic only of the impure enzyme.[208] When the enzyme was purified more than 100 fold it was found that the rate of synthesis of polymers from a reaction mixture containing only ppN was negligible. The purified enzyme can, however, synthesize polymers if an oligonucleotide primer (without a 3' phosphate) is present.

$$NpM + n(ppR) \xrightarrow{PPase} NpM(pR)_n + nP_i$$

Since the primer was incorporated into the polymer, it seemed possible to restrict the reaction to the formation of oligonucleotides by using small amounts of ppR. This has been reported by Leder, Singer, and Brimacombe, and they have been able to make just about every trinucleoside diphosphate in this manner.[121]

$$NpM + ppR \xrightarrow{PPase} NpMpR + P_i$$

An entirely different explanation of this reaction has been offered by Thach and Doty.[232,233] They have shown that under their reaction conditions the first product formed, regardless of the composition of the

initial reaction mixture, is only long polymers. Only after a relatively long incubation period does anything resembling an equilibrium mixture of oligomers form. The oligonucleotides are presumably produced by PPase acting as an exonuclease. Since PPase is incapable of degrading dimers to their component monomers, this means that the initial dinucleoside phosphate remains intact. The total reaction is as follows.

$$NpM + ppR \xrightarrow[\text{fast}]{\text{PPase}} NpM(pR)_{100\text{-}200} + P_i$$

$$\xrightarrow[\text{slow}]{\text{PPase}} NpM + NpMpR + NpMpRpR + \text{etc.}$$

The relative concentrations of the various oligomers can be adjusted by changing the stoichiometry of the reaction mixture. Thus for each synthesis the experimenter is rewarded with a series of oligonucleotides of well defined sequence. Since the dinucleoside phosphates are now commercially available, the yields of about 20% trimer relative to dimer are not at all disturbing. The only disadvantage of this method is that purified polynucleotide phosphorylase is not commercially available. Thus in principle, and apparently in practice, PPase can be used to synthesize all 64 normal trinucleoside diphosphates. But, in addition, it should be possible to synthesize many unusual trimers as well. PPase apparently readily accepts IDP and ψDP as substrates, and there is no reason why most other nucleoside diphosphates should not prove suitable.

At the time the research described in this chapter was started the elegant synthetic methods using PPase were unknown to us. Furthermore,

dinucleoside phosphates were in short supply since they had not yet become commercially available. Thus we did not seriously consider any synthetic method of preparing trimers. In addition, T1 RNAase was not yet available in large amounts. Therefore, the only practical method open to us was the degradation of RNA with pancreatic RNAase followed by separation of the oligonucleotide products. The experimental details will be described in the next section.

At the present time (November, 1965) we have successfully prepared the polynucleotide phosphorylase from M. lysodeikticus, and have used this enzyme to prepare ApCpU according to the protocol of Thach and Doty.[233] The details of this work are found in Appendix 6. They are not included in the body of the thesis because this work is far from finished. It is impressive that Thach and Doty have been able to prepare such large oligomers of well defined sequence as $(Cp)_5Ap(Up)_2U$,[233] and $(Up)_7(Ap)_3A$.[234] Suffice it to say that we feel the use of PPase is currently by far the most attractive method for preparing milligram amounts of trinucleoside diphosphates and higher oligonucleotides.

3. Preparation of Trinucleoside Diphosphates

A mixture of oligonucleotides was prepared by hydrolyzing 250 mg of "high molecular weight" yeast RNA (Worthington Biochemicals Corp.) with 5 mg of Worthington pancreatic ribonuclease (Lot #R629). This RNA is prepared according to the procedure of Crestfield, Smith, and Allen[34] and was chosen because it is believed to have a relatively long chain length, and is thought to contain only the bases A, U, C, and G. A solution of the RNA in 50 ml of distilled water was adjusted to pH 7.9

with KOH and the enzyme was added as a dry powder. The reaction was followed to completion by titration with 0.33 N KOH in a Radiometer TTT-1 Titrimeter, with intermittent incubation at 37° C. The terminal phosphates were removed by treatment for 24 hours with 0.5 mg of Worthington Bacterial Alkaline Phosphatase (Lot #6136) (EC 3.1.3.1). Then solid urea was added to bring the solution up to 7 molar.

DEAE Cellulose (Bio-rad Cellex D, 0.86 meq/gm, control #B-2276) was prepared in the following manner.[251] Fifty grams of dry resin was allowed to settle for 17 minutes at least 3 times in 4 liters of water. The fines were decanted and the water was removed by filtration. The resin was allowed to soak successively for 2 hours in .1 N HCl containing 2 N NaCl, 1 hour in 1 M $NaHCO_3$, 1 hour in 1 M Na_2CO_3, 1/2 hour in 0.5 M NaOH, and 1/4 hour in 95% EtOH, after being washed with water before each step. Then the resin was soaked in 7 M urea containing 3 M NaCl for 1/2 hour, washed with water and stored as a wet cake. B&A ACS reagent urea (lot #X139) was used throughout this work. Just prior to packing the column the resin was resuspended in 7 M urea containing 1 M NaCl. A continually stirred slurry was poured into a 1 x 70 cm column and allowed to settle. The column was washed with 1/2 liter of 3 M NaCl-7 M urea before each use. Then unbuffered 7 M urea was allowed to flow through the column until no chloride ion could be detected with $AgNO_3$. (This usually required several liters of elution.)

The above mixture of oligonucleosides was placed on the column in about 150 ml of 7 molar urea and about 1.5 liters of urea were allowed to flow through the column before starting a gradient of .05 molar NaCl per liter for 8 liters of 7 M urea-tris-chloride. This permitted a very

effective separation of oligonucleosides according to chain length.[241] It also separated the two nucleosides C and U. Typical examples of this chromatography are found in the theses of Warshaw[252] and Yolles.[264] In our hands it was found that higher resolution could be obtained if the column was run at pH 5 (acetate) instead of pH 7.8.

The trinucleoside diphosphate fraction (the fourth major peak of a column run at pH 7.8) was dialyzed in standard Visking cellulose tubing against continuously flowing water for two days, then lyophilized to dryness, dissolved in about 1/10 the original volume of water, redialyzed against water and relyophilized. The resulting white powder was stored in a dessicator until ready for use. Several unseparated mixtures of trinucleoside diphosphates prepared by the above method were given to us by Warshaw. The following separations were carried out using his fractions 12-C and 13-C.[251]

DEAE cellulose, prepared as above, was equilibrated in 7 molar urea containing 0.1 M formic acid and 1 M NaCl. This was packed as before into a 1 x 70 cm column and was washed first with 7 M urea containing 0.1 M formic acid and 3 M NaCl, and then with 7 M urea-.1 M formic acid until no chloride ion could be detected in the effluent. One of the above mixtures of trimers was dissolved in 5 ml of unbuffered 7 molar urea and placed on the column. The trinucleoside diphosphates were eluted with 7 M urea-.1 M formic acid. A Durrum peristaltic pump was used to maintain a constant flow rate. A stepwise NaCl gradient was used, as shown in Figure 1. This chromatographic procedure is very reproducible. By analogy with the chromatography of the corresponding trinucleotides by Rushizky and Sober,[190] we tentatively assumed that

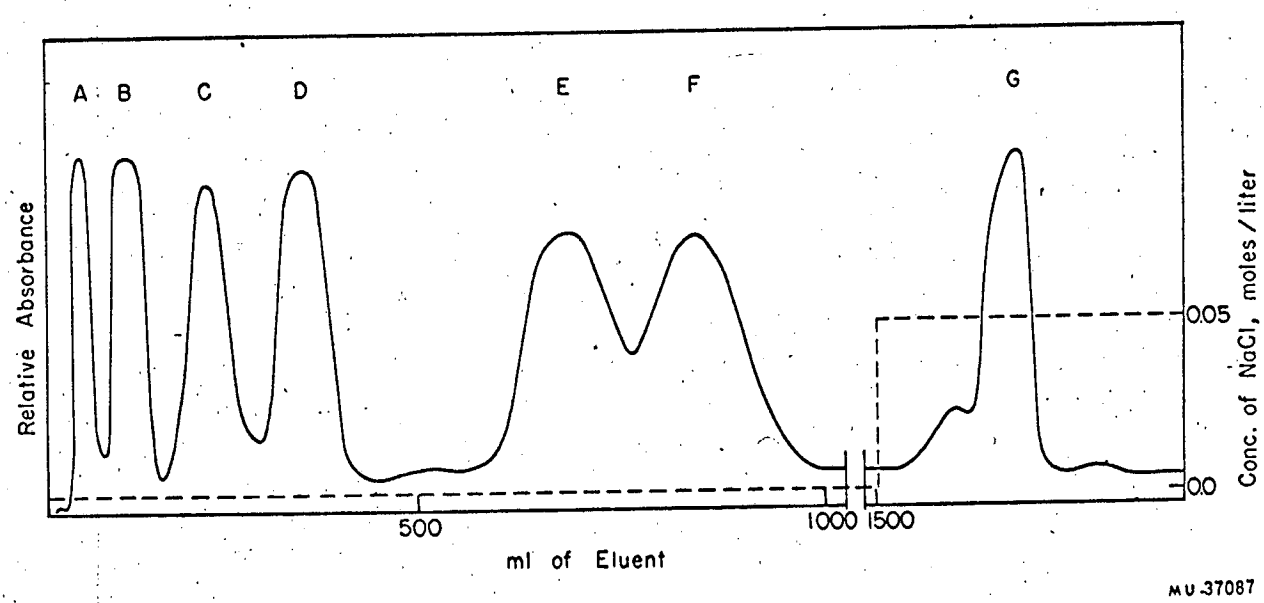


Figure 1. Chromatographic separation of eight trinucleoside diphosphates on DEAE-cellulose at pH 3 in the presence of 7 M urea: (A) ApApC; (B) ApGpC + GpApC; (C) ApApU; (D) GpGpC; (E) GpApU; (F) ApGpU; (G) GpGpU.

the peaks contained the following compounds: (A) ApApC; (B) ApGpC + GpApC; (C) ApApU; (D) GpGpC; (E) and (F) GpApU and ApGpU; and (G) GpGpU. From the absorption spectra of these peaks this assignment seemed very likely. The separated trinucleoside diphosphates were dialyzed against continuously flowing water, lyophilized, redialyzed and relyophilized. Inspection of the spectra of the pure compounds thus obtained confirmed our earlier assignment of the above chromatographic peaks. Later studies of the ORD permitted us to show that peak E is GpApU and peak F is ApGpU. We have no explanation why these two isomers separated in this system and why ApGpC and GpApC did not. Attempts made to separate the isomers ApGpC and GpApC on a Dowex-1 x2-formate column using a formic acid gradient from 0 to 1 molar in 6 liters were without success. Equally fruitless was an attempt to resolve the isomers on Whatman #1 paper using either n-propanol, water, ammonia (60:10:30, v/v/v) as the solvent, or 0.1 M phosphate buffer, pH 7.0, containing 40 gm $(NH_4)_2SO_4$ per 100 cc of solution.

The yields of trinucleoside diphosphates prepared by the above procedure were very low. Since we started with about 250 mg of RNA, which is approximately equimolar in the four bases, we would expect 6 mg of each trimer as a theoretical yield, or a total of 48 mg of trimers. After extensive dialysis, the mixture of trimers obtained by chromatography at pH 7.8 contained approximately 35 mg of material. This was chromatographed and dialyzed as described above and resulted in a total of 8 mg of separated trimers. When these were redialyzed, the following approximate yields were obtained.

	ApApC	0.5 mg
ApGpC +	GpApC	0.67 mg
	ApApU	1.0 mg
	GpGpC	0.9 mg
	GpApU	1.3 mg
	ApGpU	0.8 mg
	GpGpU	1.4 mg

Thus, we were able to recover only a total of 6.4 mg of the 48 mg of trimers theoretically contained in the RNA. It is easily seen that the greatest source of loss is due to the continuous dialysis performed on the separated trinucleoside diphosphates. In the future this step should probably be replaced by a chromatographic separation of the urea and formic acid from the trimers using a volatile buffer. The latter could then be removed by lyophilization. In spite of the low percentage yield, the actual amounts of trimers obtained were more than enough for our needs. If, however, we had chosen to use T1 hydrolysis instead of pancreatic RNAase, the yields might have been a factor of two lower, which would have caused some problems.

Because of the small amounts of material involved, it was impossible to store the trimers lyophilized since they did not form a visible powder but instead comprised a film on the walls of the vessel used for lyophilization. So, instead, they were stored as frozen solutions after the dry trimers had been leached off the walls of the containers, using between 1 and 2 ml of water. But even after one year of storage no degradation of ApGpC + GpApC stored in this way was observed by chromatography in several systems.

ApApA was obtained by hydrolysis of poly A with KOH and chromatography on Dowex-1 x 2 formate. We have used the results of S. Davis[38] and P. Wahl.[250]

4. Experimental Procedures

(a) General Methods

In all of the procedures described below we have tried to stick closely to the methods employed by Warshaw in his study of the corresponding dinucleoside phosphate properties.[252] This was done to make comparison of trimer and dimer data as unambiguous as possible. Nevertheless, there have been slight variations, and thus we shall discuss all of our methods in detail. There has been some evidence, from other laboratories, that some of the optical properties we measured are very sensitive to slight changes in pH.[245] But we have not found this. In addition, the agreement between optical data on the trinucleoside diphosphates from our laboratory and data obtained by other workers (where direct comparison is possible) has been excellent.

(b) Solvents and Solutions

Three buffers were used with nominal pH's of 1, 7, and 11.5. A 1.99 molar solution of perchloric acid was used for the pH 1 stock solution. This was diluted twofold to make a buffer of pH 1.08, ionic strength 0.1. A solution of 0.0111 molar KH_2PO_4, 0.0096 molar Na_2HPO_4, and 0.160 molar $KClO_4$ was used for the pH 7 stock solution. When diluted twofold it had a pH of 6.80 and ionic strength 0.1.[169] A solution

of 0.0063 molar NaOH and 0.20 molar $KClO_4$ was used for the pH 11.5 stock solution. When diluted twofold it had a pH of 11.42 and ionic strength 0.1. These buffers were chosen because they are relatively transparent in the UV.

Stock solutions of the trinucleoside diphosphates in distilled water were made up to approximately O.D. 2, and stored frozen. Just prior to use, the stock solutions were diluted 1:1 with the buffer stock of the particular pH desired. This was done to minimize the possibility of hydrolysis or degradation at pH 11.5 or 1, respectively. Occasionally the time dependence of spectra was checked to make sure no degradation was occuring. No time dependence was observed in any of our spectra over periods frequently as long as several hours. These time intervals are as long as trinucleoside diphosphates were ever exposed to either of these two extreme pH's.

(c) Determination of Extinction Coefficients

The extinction coefficients of the trinucleoside diphosphates were found by hydrolyzing them to nucleotides and nucleosides with Worthington Venom Phosphodiesterase and measuring the absorbance change and final absorbance for this process. The phosphodiesterase was assayed in the standard manner against bis-p-nitrophenylphosphate[262] at pH 8. At this pH our samples of enzyme contained more than 200 times more phosphodiesterase activity than phosphomonoesterase activity (as measured against p-nitrophenylphosphate). Occasional assays were run at pH's as low as about pH 7, where it was found that the enzyme was still active, although the activity was reduced at least by a factor of 5. Thus it was decided to measure extinction coefficients by hydrolysis at pH 7.3.

0.5 ml of a solution of 0.02 molar Tris-ClO_4, containing 0.02 mg per ml of freshly assayed venom phosphodiesterase was mixed with 0.5 ml of the trinucleoside diphosphate stock solution, as described below, and allowed to stand for 50 hours in a tightly sealed test tube. The absorbance was measured against a 1:1 dilution of the enzyme solution with distilled water. 0.5 ml of the same trinucleoside diphosphate stock solution was mixed with 0.5 ml of pH 7 buffer stock solution, using a Scientific Industries automatic syringe attachment fitted with a 1 ml syringe. The same syringe was used for all dilutions of a given reagent. Great care had to be taken not to trap air bubbles in the syringe. The absorbance of this trinucleoside diphosphate control was measured immediately against a 1:1 dilution of the buffer stock solution with distilled water. The extinction coefficient can be calculated by measuring the absorbance change and final absorbance after hydrolysis, if it is assumed that the hydrolysis is complete and that the products have the same extinction in a mixture as they do alone. The extinction coefficient of the trinucleoside diphosphate is given by

$$\varepsilon_{IpJpK}(\lambda) = \left(\frac{A_{IpJpK}(\lambda)}{A_I(\lambda) + A_{pJ}(\lambda) + A_{pK}(\lambda)}\right)\left(\varepsilon_I(\lambda) + \varepsilon_{pJ}(\lambda) + \varepsilon_{pK}(\lambda)\right)$$

where $A_{IpJpK}(\lambda)$ is the absorbance of the trinucleoside diphosphate, $A_I(\lambda)$ is the absorbance of the nucleoside, I, etc. The extinction coefficients used for the mononucleosides and mononucleotides were obtained from the extinction coefficients given for the long wavelength maxima in the Pabst Laboratories Circular OR-10:[166] for pH 7, $\varepsilon_{pA}^{259} = \varepsilon_A^{259} = 1.54 \times 10^4$, $\varepsilon_{pC}^{271} = 0.90 \times 10^4$, $\varepsilon_{pU}^{262} = 1.00 \times 10^4$ and $\varepsilon_{pG}^{252} = \varepsilon_G^{252} = 1.37 \times 10^4$. We

assume that the extinction coefficients of the monomers at pH 6.8 are the same as at pH 7.3. The extinction coefficients shown in Table I are the average of three determinations. In almost all cases the deviations of the three cases were of the order of 1%.

The oscillator strength of an absorption band is defined by the following equation,[237]

$$f = 4.318 \times 10^{-9} \int \varepsilon(\nu) d\nu$$

where the integration is carried out over the entire absorption band. If the band is not well resolved, as in the case of all trinucleoside diphosphates, dinucleoside phosphates and mononucleosides, one has to choose among several approximate procedures. The unresolved bands can be mathematically decomposed using a consistant rule for choosing the shape of the bands. This has been done by Bush for the 4 monomers,[25] but would be much more difficult to do for the trimers which have even less well resolved bands. A second possibility would be to use the minimum of each spectrum as a cut off for the integration, but this is totally unsatisfactory in the case of trimers and dimers whose absorption is the composite of many bands. The third possibility, which we have used, is to pick an arbitrary wavelength as a short wavelength cut off for the integration. This was chosen to be as near the minimum of as many spectra as possible. Oscillator strengths were obtained by integrating the extinction coefficients from 350 mμ to 232.5 mμ for data at pH 11.5 and pH 1, and from 350 mμ to 230 mμ for pH 7 spectra. A computer program written by Warshaw was used to perform these integrations numerically. Details of this program, as well as evidence from dinucleoside phosphate spectra which show the effect of the choice of cut off

on the oscillator strength obtained, are given in the thesis of Warshaw.[252] To permit comparison with our data, oscillator strengths of dinucleoside phosphates and mononucleotides were occasionally recalculated to conform to our choice of boundaries for the integration. Except for pG which was obtained from Pabst Circular OR-10,[166] all monomer and dimer data used for computing oscillator strengths were obtained from Warshaw.[251]

(d) Optical Measurements

All UV spectra described in this chapter were taken on a Cary 15 Spectrophotometer at room temperature. The ORD was measured on a Cary 60 Spectropolarimeter using a pen period of 3 seconds and a slit program chosen to keep the spectral bandwidth at less than 1 mμ. Occasional samples were run in duplicate to check the reproducibility of the measurements. A 1 cm cell with a total volume of 0.7 ml was used for all measurements.

(e) Treatment of ORD and Absorption Data

The ORD of a 1:1 dilution of the appropriate buffer blank with water was measured just prior to or just after the measurement of the ORD of each of the trinucleoside diphosphates. This baseline was subtracted point by point at 5 mμ intervals from the ORD of the buffered trinucleoside diphosphate solution, after the noise in both readings had been averaged by eye. The ORD is expressed as molar rotation per residue

$$[\Phi] \equiv \frac{[\Phi']}{n} = \frac{100\theta}{c\ell}$$

where $[\Phi']$ is the molar rotation, n is the number of base residues per

mole, ℓ is the path length in cm, θ is the measured rotation in degrees, and c is the molar concentration of base residues.

The absorption spectra were digitized by hand at 2.5 mμ intervals, and corrections for the baseline were made. Extinction coefficients are usually expressed in this work as molar extinction per residue, but in the illustrations for this chapter the data are plotted as molar extinction per oligomer.

5. Evidence for Base Stacking

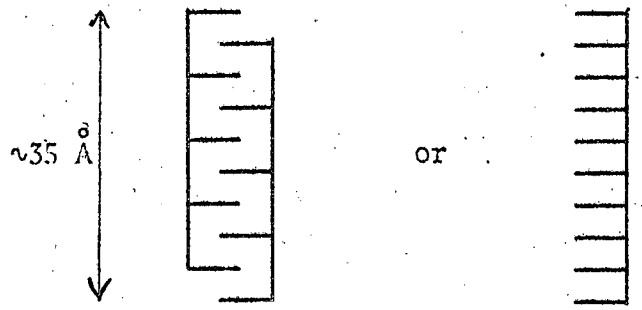
Ever since the discovery that the base composition of most types of RNA does not lend itself to the formation of a regular double strand helix with complementary base pairs, most speculation on the conformation of RNA has revolved around substantially less ordered structures.[217] In the extreme case, one could suppose that an RNA strand in solution is a coil consistent with the constraints imposed by the electrostatic free energy. However, under the normal conditions in which RNA is studied by the physical chemist (pH 7, 0.1 M salt is a rough approximation) a random coil is not consistent with most experimental results. Alternative structures which have received much favor are built on the postulate that RNA strands try to form as many intramolecular complementary base pairs as possible. This type of structure will be discussed in detail in Chapter IV. Recent evidence from a wide variety of experimental measurements has led to the concept of a single strand stacked helix as a common RNA conformation. In this section, we shall review some of the data which has convinced us that the bases of single strand oligonucleotides and polynucleotides often are found in a stacked conformation.

One of the most direct pieces of evidence for base stacking has come from the study of the intermolecular interaction of individual nucleosides or bases. From vapor pressure studies of aqueous solutions it was found that 6-methylpurine associates more than purine, and 5-bromouridine associates more than uridine. This cannot be explained in terms of hydrogen bonding, and Ts'o and Chan suggest that the difference in free energy of association is due to hydrophobic interactions.[242] More convincing evidence comes from a study of the proton NMR of aqueous solutions of purine and 6-methylpurine. In these two compounds, the proton chemical shifts relative to chloroform move to higher field with increasing concentration. This is what would be expected if dimers or higher aggregates are formed in which the base planes are stacked upon one another. The effect would be of opposite sign if the aggregates contained bases in the same plane.[31]

A third line of evidence which suggests base stacking in solutions of monomers comes from the unusual properties of concentrated solutions of guanosine 3' or 5' phosphates. It has long been known that these solutions readily form gels. An X-ray study of these gels by Gellert, Lipsett, and Davies suggests that the structural element of the gel consists of plates of 4 G's hydrogen-bonded together.[70] These plates are then stacked upon one another in a helical array to form the high molecular weight aggregate. Unlike the previous two cases, hydrogen bonds do play a role here, but the stacking interaction is also of importance.

Stacked bases in nucleic acid polymers have been known for quite some time from the X-ray structure of DNA fibers. They were at first thought to be just a result of the constraint of maximal hydrogen bond

formation rather than an independent source of stability. However, recent evidence suggests that the stacked bases remain even when there is no possibility of interstrand hydrogen bonding. Small angle X-ray scattering on solutions of DNA in glycol or DNA which has been treated with formaldehyde has been performed by Luzatti and his collaborators.[137] Formaldehyde reacts with the bases and will effectively block most possibilities of hydrogen bonding. Yet in both this case and in glycol solutions of unreacted polymer, DNA shows a mass per unit length of between 3.2 and 3.5 angstroms per nucleotide residue. This structure is consistent with either a single strand helix with stacked bases or a double strand helix with intercalated bases. These two structures are shown schematically below.



The small angle scattering of poly A at pH's from 6 to 7.2 yields very similar results. There is one nucleotide every 3.5 angstrom.[137,260] Luzzati concludes that the structure of the neutral form of poly A is thus probably very similar to one of the DNA structures shown above, but he is unable to decide between the two.

However, the recent ORD data on poly A obtained by Holcomb and Tinoco,[95] when used in conjunction with the ORD study of ApA by Warshaw, Bush and Tinoco,[253] permit us to decide between the two possibilities.

The ORD spectra of poly A and ApA are very similar in shape, although different in magnitude, and thus we can say that the polymer and oligomer have very similar conformations. These curves are reproduced in Figure 2. The ORD of ApA has been calculated by Bush,[25] using the experimentally determined transition moment of the 260 mμ absorption band. He assumed that the dinucleoside phosphate has stacked bases in the geometry they would have if they were nearest neighbors on a single strand of double strand DNA. The agreement between the experimental and theoretical ORD curves, reproduced in Figure 3, is very encouraging, and strongly suggests that the bases in ApA are stacked. In turn, this suggests that poly A is a single strand stacked helix.

Additional calculations by Bush are not all so encouraging. There is still good agreement between the experimental and theoretical ORD curves of ApU and UpA. When the comparison is made between oligomers that contain either G or C, there is much less agreement. This is certainly due, in part, to the fact that experimental transition moments are not available for these compounds. Until more trustworthy calculations are available we must be content in interpreting the ORD of oligo and polynucleotides in a more qualitative manner. The assumption which has usually been made is that an ordered structure exists when the ORD of the oligomer or polymer is substantially different from the sum of the optical rotation of its componont monomers. That this is certainly the case for poly A and ApA can be seen in Figure 2. Additional data from Warshaw's thesis[252] shows that at pH 7 most dinucleoside phosphates have a molar rotation which is quite different from the sum of their two monomers. Thus, in analogy with poly A and ApA, the assump-

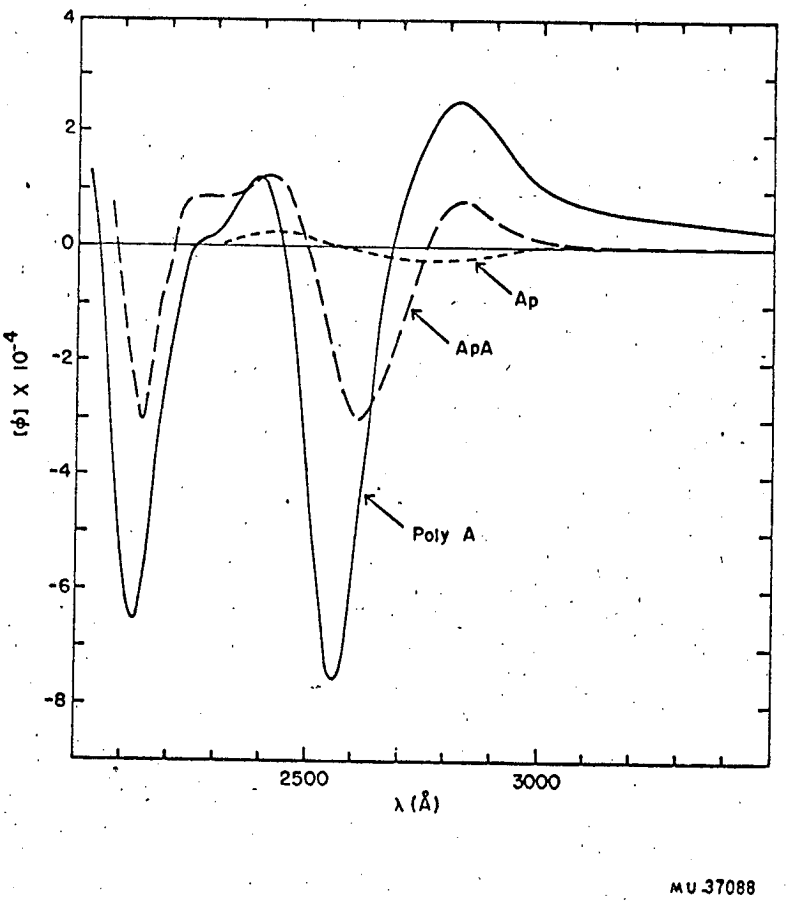


Figure 2. Optical rotatory dispersion of poly A, ApA, and Ap at neutral pH. Taken from the work of Holcomb and Tinoco.[95]

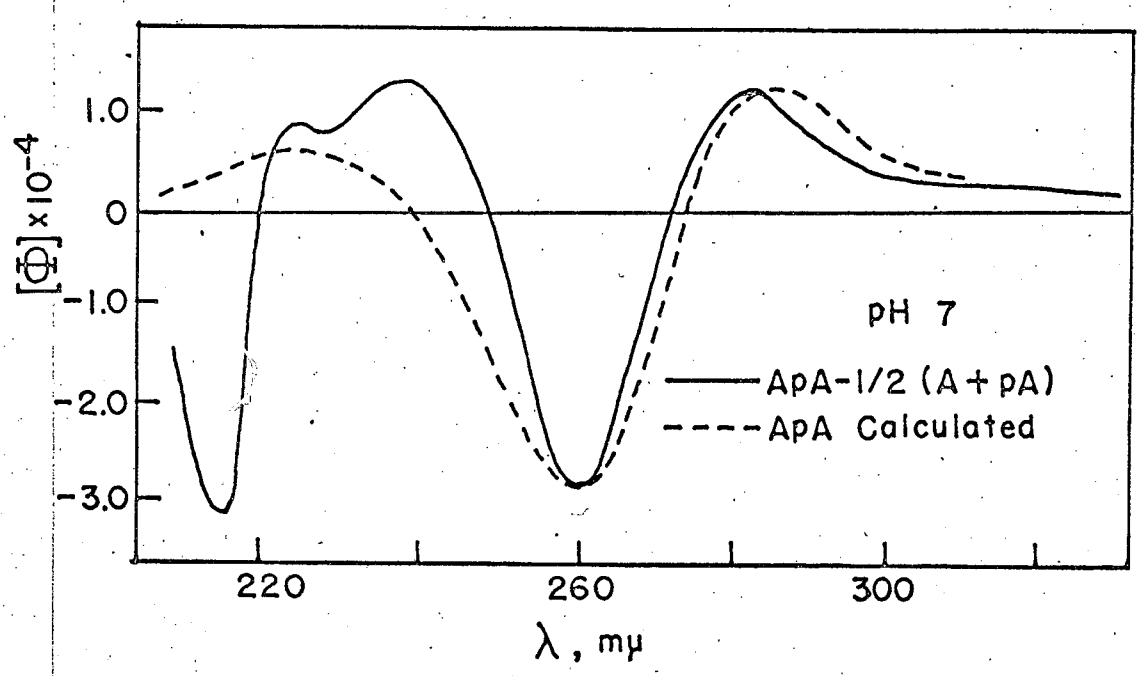


Figure 3. Optical rotatory dispersion which results from base-base interactions in the dinucleoside phosphate ApA. Redrawn from the thesis of Bush.[25]

———— ApA - 1/2(A+pA) experimental

- - - - - -ApA calculated for two bases held rigid in the geometry of DNA. Magnitude fit at 260 mμ.

tion was made that these oligomers have stacked bases. While this assumption is very likely, it is not yet possible to draw a conclusion from ORD data alone that the bases in dimers other than ApA are stacked. However, other optical data from many other sources tends to confirm these assumptions.

It is known that the temperature dependence of the ORD or absorbance of a highly ordered double strand structure like DNA shows a very sharp transition. This is attributed to strand separation. However, the transition from a more structured to less ordered single strand polynucleotide is extected to be much broader.[74] We can use this as a criterion to distinguish between single and double strand RNA conformations. At pH's lower than 6, Luzzati and his coworkers have found that poly A has 2 residues per 3.5 angstroms.[137] This is consistent with the double strand structure suggested by X-ray fiber studies on low pH poly A. Holcomb and Tinoco have found that at pH's below 5 the ORD of poly A undergoes a sharp change with increasing temperature.[95] In contrast, at pH 7 the temperature transition of the ORD of poly A is very broad. This strongly suggests that the neutral form of poly A is a single strand stacked helix and not a double strand with intercalated bases. Similar experiments have been reported by several other workers.[116,195,19]

Recent data on the ORD of poly C reinforce this result.[54] The temperature dependence of the neutral form is very broad. For an additional comparison, the poly C was reacted with formaldehyde. This prevents the formation of hydrogen bonds. Yet the temperature dependence of the ORD was almost the same before and after formaldehyde reaction.

Thus poly C is a single strand helix at pH 7. Reaction with formaldehyde changes the absorption spectrum of RNA considerably, and thus makes it difficult to arrive at quantitative conclusions. But Fasman, Lindblow and Seaman have found that the ORD of formaldehyde treated sRNA is substantially different from the ORD of the monomers reacted with formaldehyde.[55] This again suggests that RNA is capable of existing as a single strand stacked helix. More evidence in this vein comes from the work of Van Holde, Brahms, and Michelson.[246] They prepared poly 6-N-hydroxyethyladenylic acid. This shows very strong circular dichromism, unlike the monomer, even though it is impossible to form any hydrogen bonded base pairs. Thus there is a large amount of evidence that polynucleotides can have a single strand ordered conformation. All of this evidence is consistent with the presence of intramolecular stacking of bases.

Some of the most pointed evidence for intramolecular base stacking has been obtained through the study of oligonucleotides. We have already discussed the work of Warshaw, Bush and Tinoco on the ORD of ApA. Similar results have been obtained by other workers.[147,246] These workers have concluded that the bases of ApA are stacked. Other evidence comes from the study of the hypochromicity of oligonucleotides and their synthetic analogues. The hypochromicity[141] at a given wavelength is defined by the following equation,

$$H(\lambda) = 1 - \varepsilon_p(\lambda) \; / \; \sum_m \varepsilon_m(\lambda)$$

where ε_p is the molar extinction coefficient of the polymer, and ε_m are the monomer extinction coefficients. The hypochromicity is usually re-

ported at the maximum wavelength or at 260 mμ for nucleic acid materials. The reader should be cautioned that some workers report their results as hyperchromicity instead. This is given by

$$H'(\lambda) = \sum_m \epsilon_m(\lambda) / \epsilon_p(\lambda) - 1$$

The hyperchromicity and hypochromicity are related by the following equations,

$$H'(\lambda) = H(\lambda) + \left(\sum_m \epsilon_m(\lambda) - \epsilon_p(\lambda) \right)^2 / \epsilon_p(\lambda) \sum_m \epsilon_m(\lambda)$$

or

$$H(\lambda) = H'(\lambda) / (1 + H'(\lambda)) \quad .$$

Thus we can see that to first order in $\epsilon_m - \epsilon_p$ the two expressions are the same, but a small second order correction is necessary. Further complications ensue when hypochromicity is reported as acid or basic hyperchromicity. These are the fractional increaseses in absorbance when the pH is changed from neutrality to pH 1 or to a strongly alkaline pH. These two hyperchromicities are extremely misleading, and the use of this quantity should be avoided. Comparisons between hypochromicity as defined above and acid or basic hyperchromicity are not necessarily meaningful.

The hypochromicity at a given wavelength is related to the hypochromism of the absorption band which is defined by the expression

$$H = 1 - f_p / \sum_m f_m$$

where f_p and f_m are respectively the oscillator strengths of the polymer and monomers. Thus we can see that the hypochromism and hypochromicity have similar form. If the absorption band does not change shape or

shift in going from the monomers to the polymer, but merely increases in magnitude; then the numerical values of hyperchromicity and hypochromism should be similar.

Theoretical work by Tinoco has established that hypochromism arises from interaction between nucleic acid bases which are stacked vertically.[235,41] Interaction between bases in the same plane will result in hyperchromism (or hypochromism with a negative sign). If we make the approximate connection between hypochromism and hypochromicity, we are now able to interpret much of the optical data compiled by Michelson.[151] The phenomenon of hypochromicity has been known for a long time in nucleic acid materials. Poly A, poly C, poly T, and poly I are all thought to be single stranded at sufficiently low concentrations at neutral pH's, and yet all of these molecules show very substantial hypochromicity. Thus the important conformations which contribute to their structure must have stacked bases. All of the dimers, trimers and tetramers whose spectra are summarized by Michelson show appreciable hypochromicity. The 2'-5' synthetic oligomers seem to be more hypochromic than the natural 3'-5' isomers. In addition, even 5'-5' dinucleoside diphosphates show substantial hypochromicity. In these compounds, shown schematically below, the bases are considerably further away from each other than in 3'-5' dimers. Yet the bases still seem to prefer to stack.

Base-ribose-5'-phosphate-phosphate-5'-ribose-base.

Other analogues of oligonucleotides which are hypochromic compared to their monomers are α,ω,di-9-purinyl ethanes or hexanes.[151] In these

compounds the idea of intramolecular hydrogen bonds holding the conformation together can certainly be ruled out.

One of the difficulties with studying oligomers as models for polymers is that the former are not constrained to form any sort of a regular structure. Thus, for example, a trimer could have a conformation with stacked bases where base three has been intercalated between one and two instead of the more regular order of one on top of two on top of three. In the polymer, such a structure would be unlikely for steric reasons. Thus it is comforting to see that data on oligomers does support the idea of a linear stacked trinucleoside diphosphate. For example, the trimer GpUpA shows appreciable hypochromicity, while the analogue GpdiHUpA has no hypochromicity. If, however, the diHU is on the end of the trimer the hypochromicity reappears.[151]

Both the hypochromism and ORD of dinucleoside phosphates have been studied by Warshaw as a function of pH.[252,254] When the pH is lowered from 7 to around 1, both A and C acquire a positive charge. Thus in oligomers like ApA and ApC, the bases would be expected to repel one another away from the stacked conformation. This is corroborated by the optical evidence which shows that at pH 1 these oligomers have the same absorption spectra and ORD as the sum of their monomers.

We have shown that evidence from a variety of sources indicates that the bases in single strand polynucleotides and oligonucleotides often have a stacked conformation. Several explanations have been offered to account for this. Hydrophobic forces would be expected to favor a conformation in which the aromatic bases are exposed to the solvent as little as possible. This occurs in the stacked conformation. A

more quantitative explanation of this has been put forth by Sinanoğlu and Abdulnur.[204] They show that if water is considered as a macroscopic phase, the stacked conformation would minimize the surface free energy of the aqueous shell around the bases. The stacked conformation also maximizes the London attraction between the bases.[40] Quantitative determinations of the stacking energy are now being determined in our laboratory.[37]

Experiments by Davis have shown that when oligonucleotides are dissolved in solvents other than water the stacked conformation is destroyed. Thus the ORD of ApA in methanol, reproduced in Figure 4, or ApAp in ethanol, becomes like that of the monomers in these solvents. Additional experiments have been reported using trimethyl phosphate as the solvent. These results are attributed to disruption of the hydrophobic forces in solvents other than water.[38] Additional evidence comes from the study of the photo-induced dimerization of thymine bases in TpT. Wacker and Lodemann have found that the extent of the photoreaction per unit dose of radiation strongly depends on the solvent.[249] The stacked conformation is thought to favor the photodimerization. Thus it is very encouraging that the extent of photoreaction correlates very well over a large series of solvents with the stacking energy calculated by Sinanoğlu and Abdulnur.

6. Conformation of Dinucleoside Phosphates

Up till now we have used the term base stacking very loosely. Stacking means that the base planes of nucleotides are approximately parallel and are nearly as close atop one another as the Van der Waals radii will

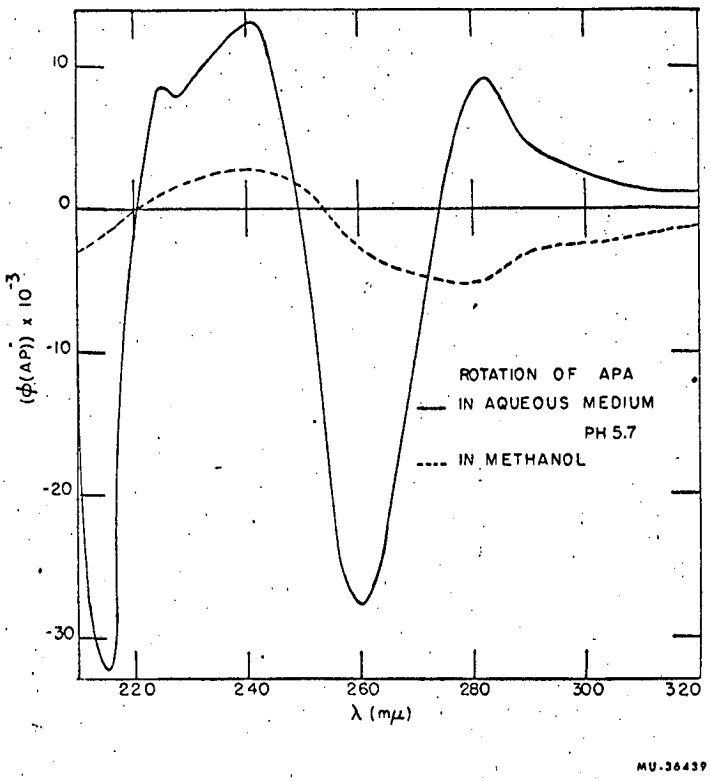
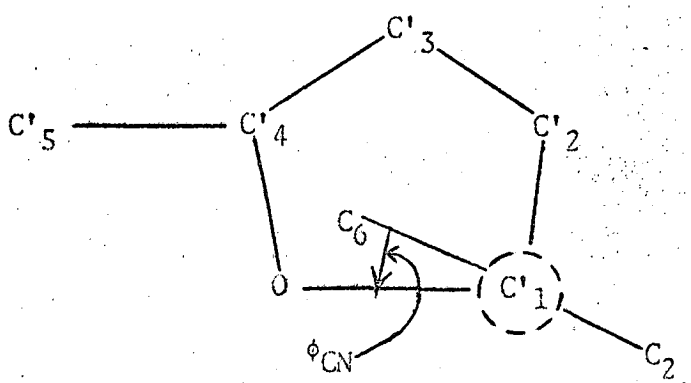


Figure 4. Optical rotatory dispersion of ApA in methanol and in water. Taken from the thesis of Davis.[38]

permit. Using this definition, it is apparent that there are many different possible dinucleoside phosphate conformations which have stacked bases. Some of these differ from one another only by small rotations around single bonds, while others involve simultaneous differences in many degrees of freedom.

There are at least seven degrees of freedom which are partially constrained when the two bases of a dinucleoside phosphate stack. These are rotation around the glycosidic linkages (C-N bond) of both nucleosides, rotation around both phosphate-oxygen single bonds, rotation around the C 3' carbon oxygen-bond and C 4', C 5' carbon-carbon bond of the 5' linked nucleoside. In addition, several other degrees of freedom involving the ring puckering of the furanose rings may be involved in stacking.[86] Thus base stacking is a fairly complicated conformational change. Only two of the degrees of freedom have been studied in any detail. This is the rotation of the bases around the glycosidic CN bond. It is described by the torsional angle, ϕ_{CN}, discussed by Donohue and Trueblood.[45] This is the angle defined by the intersection of the trace (perpendicular to the CN bond) of the plane of the base with the projection of the C'_1-O bond of the furanose ring into a plane perpendicular to the CN bond. The angle is viewed along the CN bond, and is measured in a counterclockwise direction. This angle is illustrated below for $\phi_{CN} \cong -30°$.



Donohue and Trueblood have shown from models that there are two ranges of torsional angle which seem to have much less non-bonded interaction than the others. Thus these would be expected to be the low energy conformations. These two ranges of ϕ_{CN} are defined as anti ($\phi_{CN} \sim -30°$) and syn ($\phi_{CN} \sim +150°$). Recent work by Haschemeyer and Rich[86] has quantitatively surveyed the intramolecular Van der Waals contacts of non-bonded atoms in all of the nucleotide material for which there is accurate X-ray data. All of the 15 crystal structures surveyed find the bases in the anti conformation except for the structure of deoxyguanosine in the 5-bromodeoxycytidine-deoxyguanosine complex, which is syn. This is the only crystal study yet reported for a guanine containing nucleotide, so that the generality of this phenomenon is unknown. But Haschemeyer and Rich estimate that all of the purines are capable of existing in both syn and anti conformations, and even uridine and cytidine may have important contributions from the syn conformation. Thus it is impossible to exclude the syn conformation from consideration.

The five degrees of freedom involved in the phosphate ribose backbone can arbitrarily be lumped together. Taken as a whole, they can lead to either right- or left-handed helical backbones. Each of these classes can, in principle, have many variations, but our work with Courtauld models suggests that all of the stacked conformations we can build seem to have fairly rigid backbone conformations. To change the conformation of the backbone, it is usually necessary first to unstack the bases. With the above considerations we can see that there are a total of at least eight distinct possible conformations for a stacked dinucleoside phosphate. Each base can be either syn or anti, and the

backbone can be either right- or left-handed. We have tested all of these conformations by building Courtauld models. They show that all of the eight conformations are geometrically possible. Furthermore, the range of strained angles and bonds seems to be close to the same for all of the stacked conformations. The degree of horizontal overlap does vary from conformation to conformation, but in no predictable way; and in all eight conformations, the bases can easily come as close as the Van der Waals radii of the πcloud of the aromatic rings.

Rough drawings of four of the eight approximate dinucleoside phosphate conformations of ApG are shown in Figure 5. The four conformations not shown are qualitatively similar in appearance. It can be seen that the angles between the bases, as measured by the angle between the two glycosidic CN bonds, are different in the four cases; but since a wide range of angles is possible for each of these conformations, no immediate conclusion can be drawn from this. Photographs of the Courtauld models for ApA are shown in Figure 6. A typical conformation with unstacked bases is shown in Figure 6a. Note that the bases can get very far away from one another when the sugar phosphate chain is extended. The right handed stacked conformation with both bases anti is shown in Figure 6b. For comparison, a photograph of the right handed conformation in which both bases are syn was published by Van Holde, Brahms, and Michelson.[246] It seems unlikely to us that ORD is a sensitive enough method to decide which of the eight possible conformations of a dinucleoside phosphate actually exists in solution. It may be very helpful though in narrowing down the choices. Such an approach would probably have to be accompanied by a study of suitable derivatives in which some of the conformations are impossible.

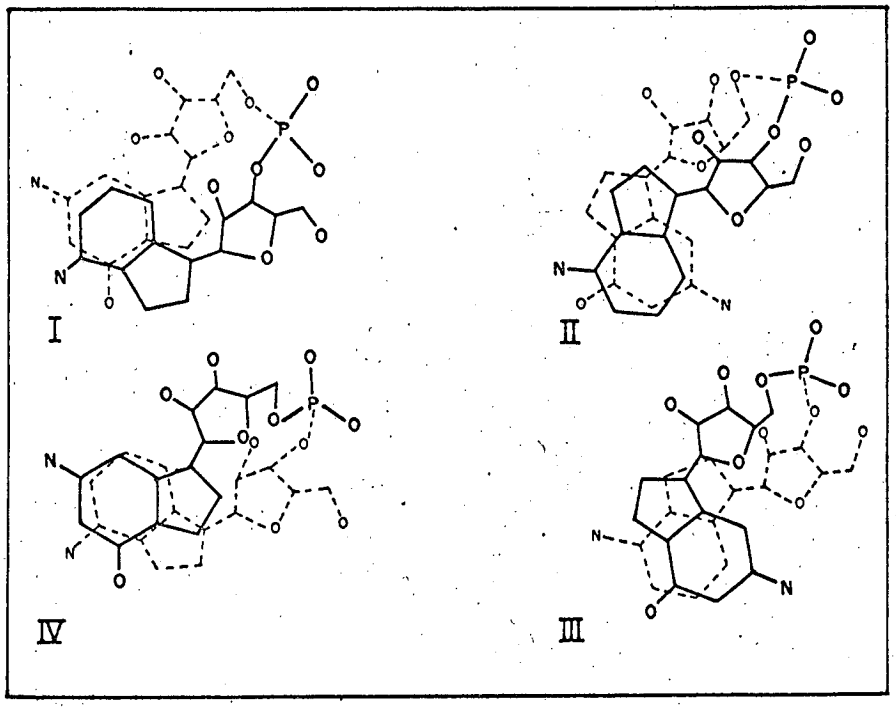


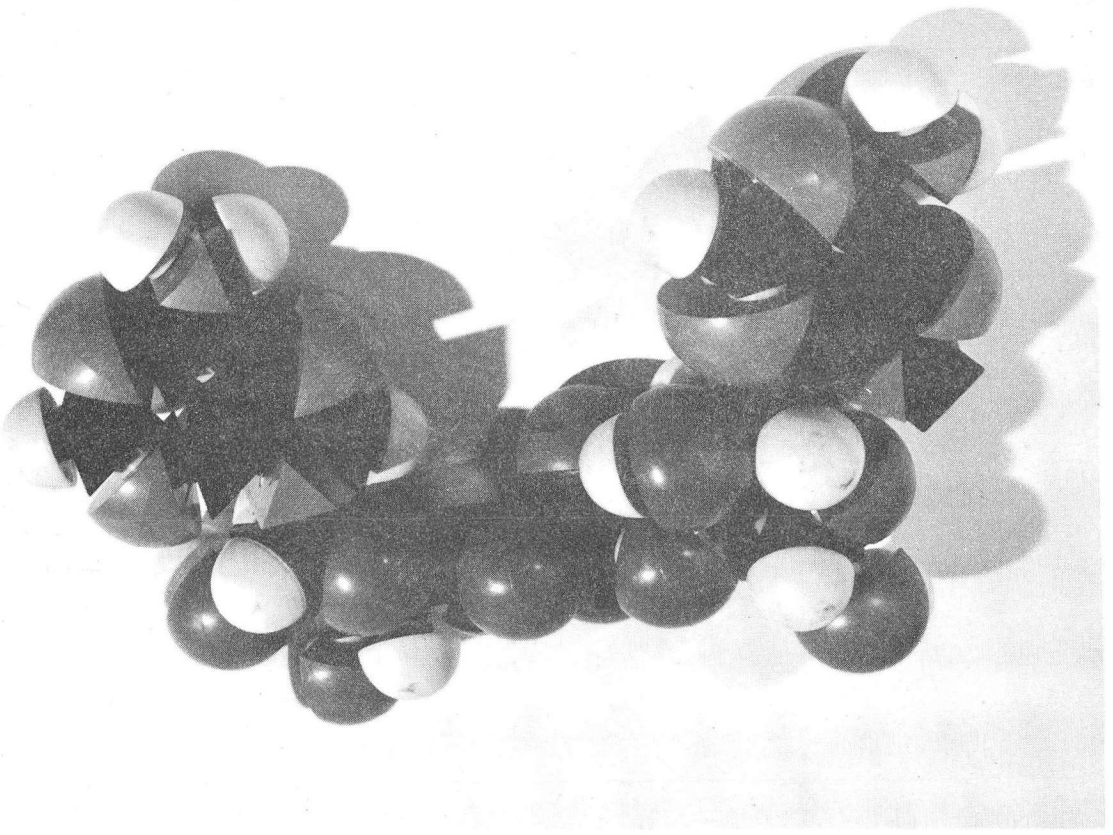
Figure 5. Four of the possible stacked conformations of ApG.

I. Left hand helix, both bases anti to ribose.

II. Left hand helix, both bases syn to ribose.

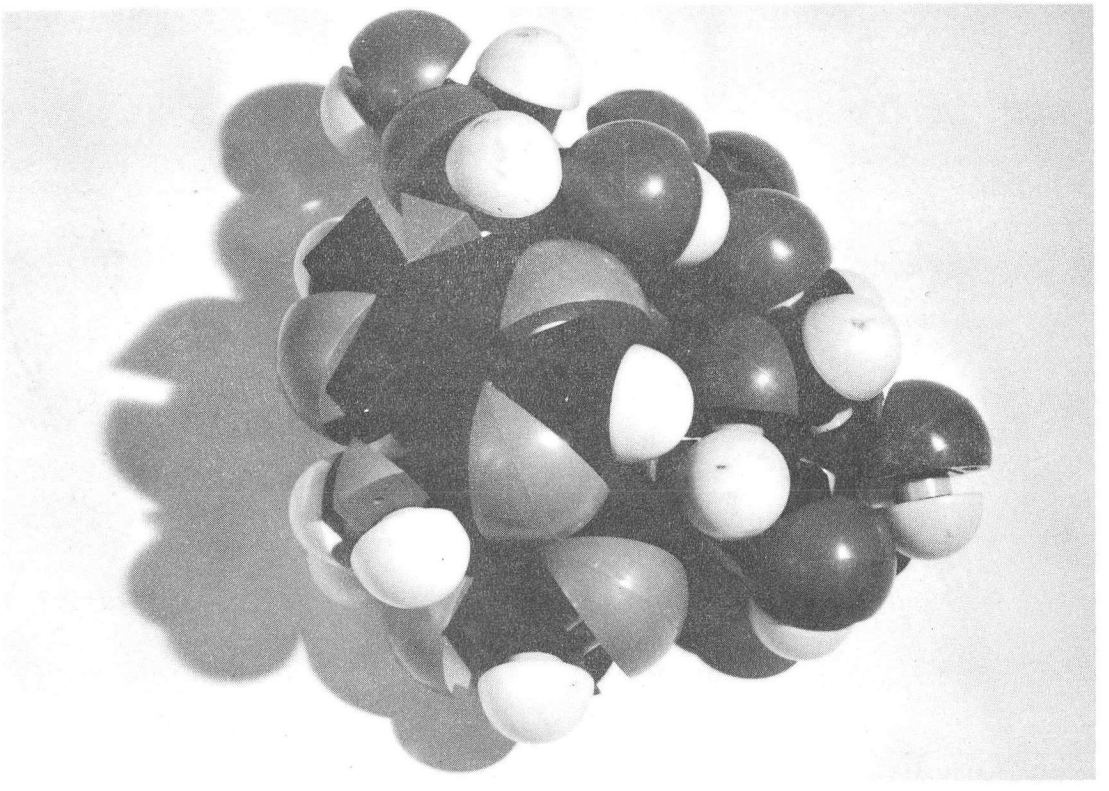
III. Right hand helix, both bases syn to ribose.

IV. Right hand helix, both bases anti to ribose.



ZN-5375

Fig. 6. Photographs of Courtauld models of ApA.
(a) Typical unstacked conformation.



ZN-5374

Fig. 6. Photographs of Courtauld models of ApA.
(b) Stacked conformation, right hand helix, both bases anti to ribose.

To resolve the problem of which dinucleoside phosphate conformations exist in solution, more powerful methods will have to be employed. NMR has shown great promise in the past in determining qualitatively, and in some cases quantitatively, the equilibrium population of rotamers around a single bond. From temperature dependence studies, even the relative energies of various rotamers can be determined.[160] However, because of the large number of protons involved in the conformational changes of a dinucleoside phosphate, methods may have to be developed which will produce a variety of deuterated derivatives.

Another method of approach which should prove useful is X-ray diffraction. Thus far, the crystal structure of only one dinucleoside phosphate has been reported.[203] Unfortunately, it was the 2'-5' isomer of ApU instead of the natural 3'-5' isomer, but the results are still of interest even though no direct comparison can be made. Both bases were found to be in the anti conformation, though the torsional angle of U was an extreme value. The two bases are stacked, and the base planes are very nearly parallel, deviating by only 15°. The distance of closest approach between the bases is 3.4 angstroms. As in all X-ray crystal studies, we are now faced with the difficult extrapolation from the crystal with all of its lattice forces to the solution where hydrophobic interactions will be more important. At the present there is no way to predict the solution conformation given that in the crystal. For want of a better assumption, we shall assume in the remainder of this work that the right hand anti-anti conformation is the favored conformation in solution. This is, however, only a working hypothesis.

7. Experimental Results--ORD

The ORD of seven trinucleoside diphosphates at three pH's is shown in Figures 7 through 21. From the experimental curves it is seen that ORD is an excellent method for identifying the distinguishing the trinucleoside diphosphates we have studied. The calculated curves, also shown in most of these figures, will be discussed later. Values for the molar rotation per residue of selected peaks and troughs at three pH's may be found in Table I. The ORD curves at pH 7 differ both in magnitude and shape for each of the trinucleoside diphosphates studied. These results, as well as the studies of 15 dinucleoside phosphates,[252,254] show that differences in ORD among oligonucleosides are much more marked than differences in their UV spectra.

The major feature of the ORD of ApApA, ApApC, ApApU, GpApU, and ApGpU at pH 7 is a double Cotton effect centered for the first four compounds at about 260 mμ and for the last at 270 mμ (see Figures 11-15). Similar ORD curves have been observed for poly A,[95] poly U, and sRNA,[116] ApA, ApU, ApC,[254] and many other dinucleoside phosphates.[252] This double Cotton effect arises from interaction between neighboring bases,[236] and suggests that these molecules contain stacked bases.[253] One may question whether the strong double Cotton effect observed for dinucleoside phosphates and trinucleoside diphosphates arises solely from intramolecular interactions. Experiments discussed in the next chapter show that there is no concentration dependence of the ORD of two dinucleoside phosphates. In addition, unpublished work by Vournakis, Scheraga, Rushizky, and Sober indicates that the ORD of ApApCp is concentration independent over the range of concentrations usually employed in optical

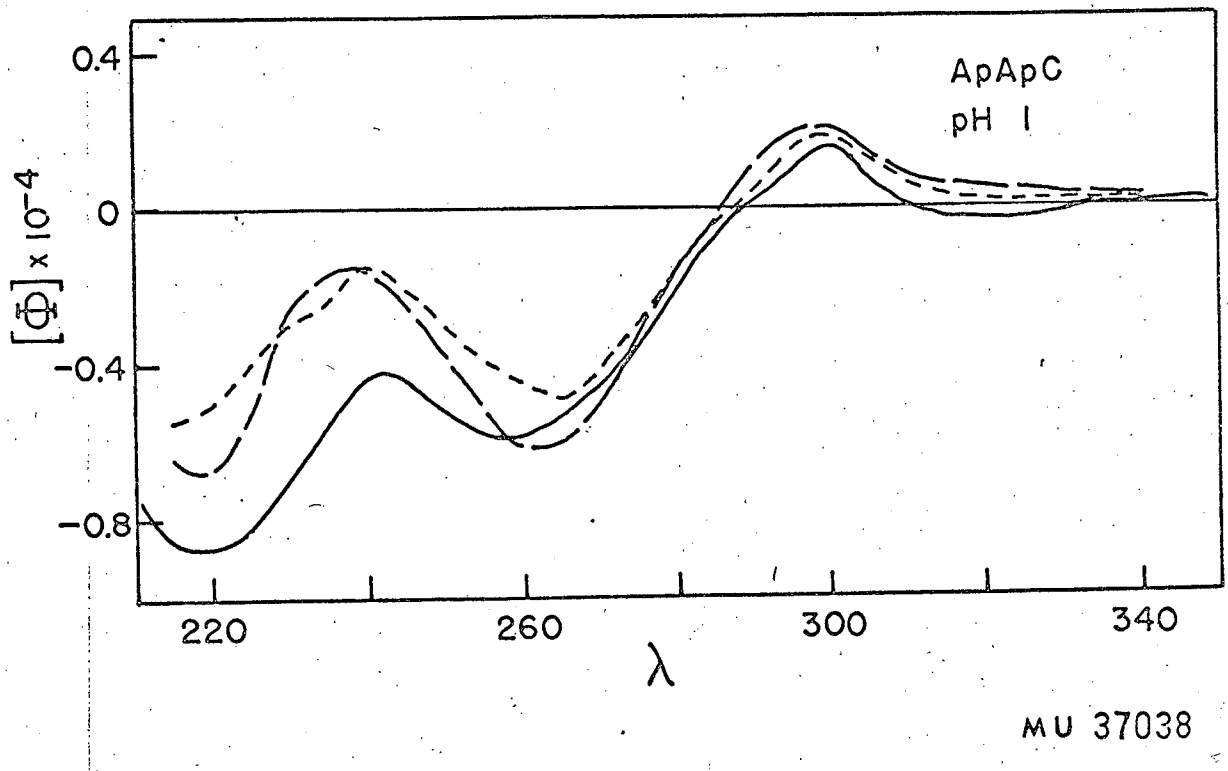


Figure 7. Optical rotatory dispersion of ApApC at pH 1.

——— experimental

— — — nearest neighbor calculation

- - - - - sum of monomers

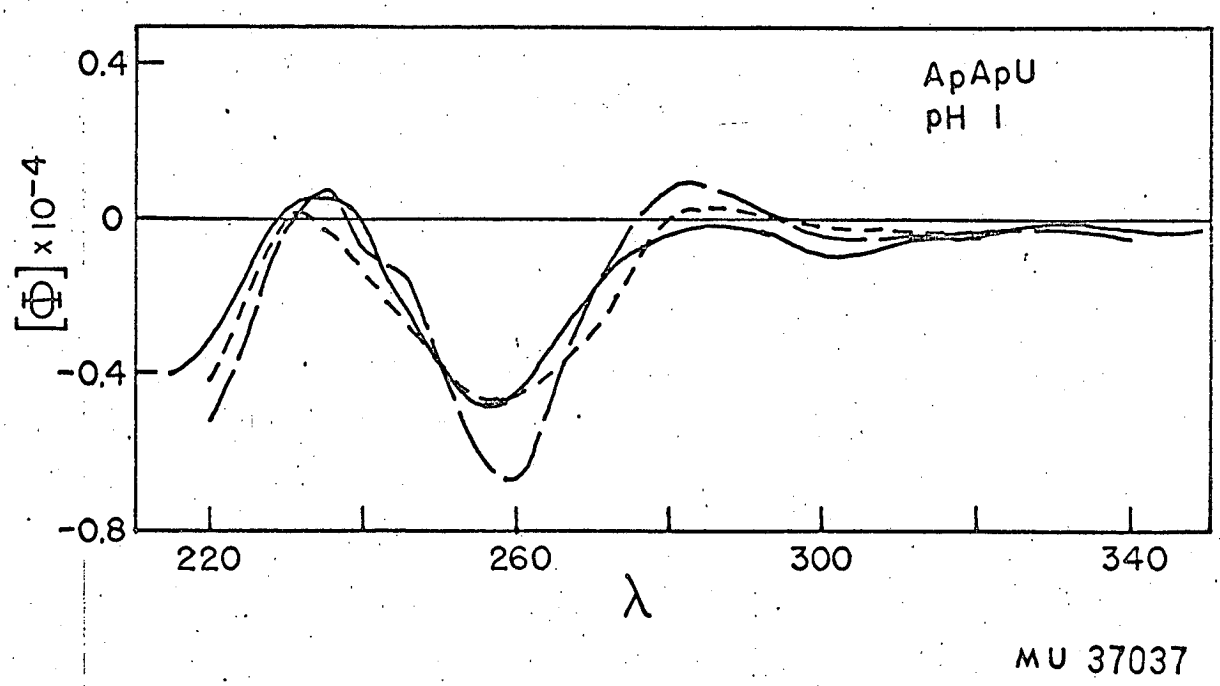


Figure 8. Optical rotatory dispersion of ApApU at pH 1.

——— experimental

— — — nearest neighbor calculation

- - - - - sum of monomers

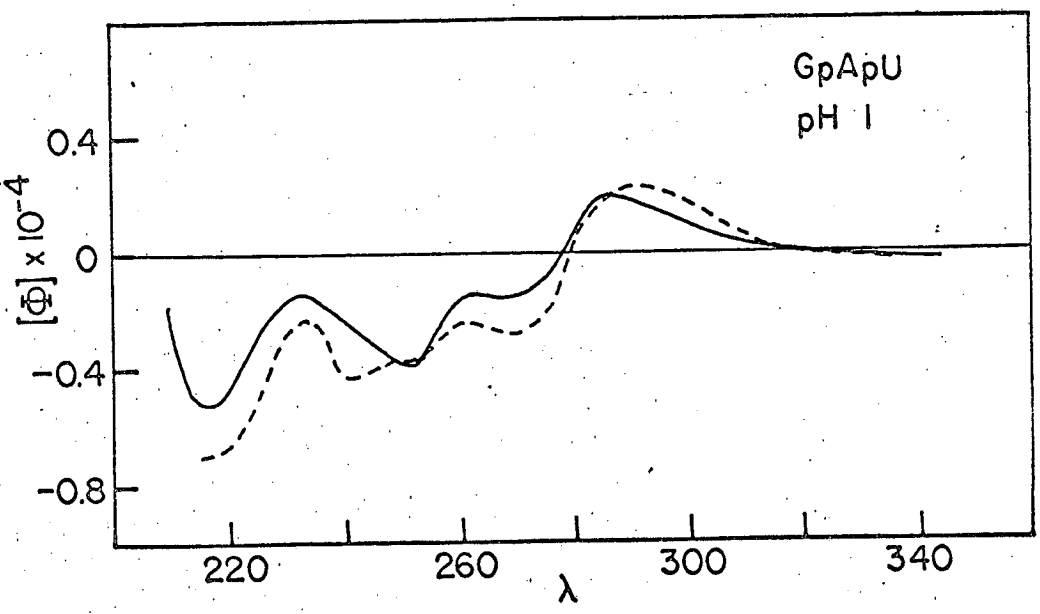


Figure 9. Optical rotatory dispersion of GpApU at pH 1.

——— experimental

– – – – – nearest neighbor calculation

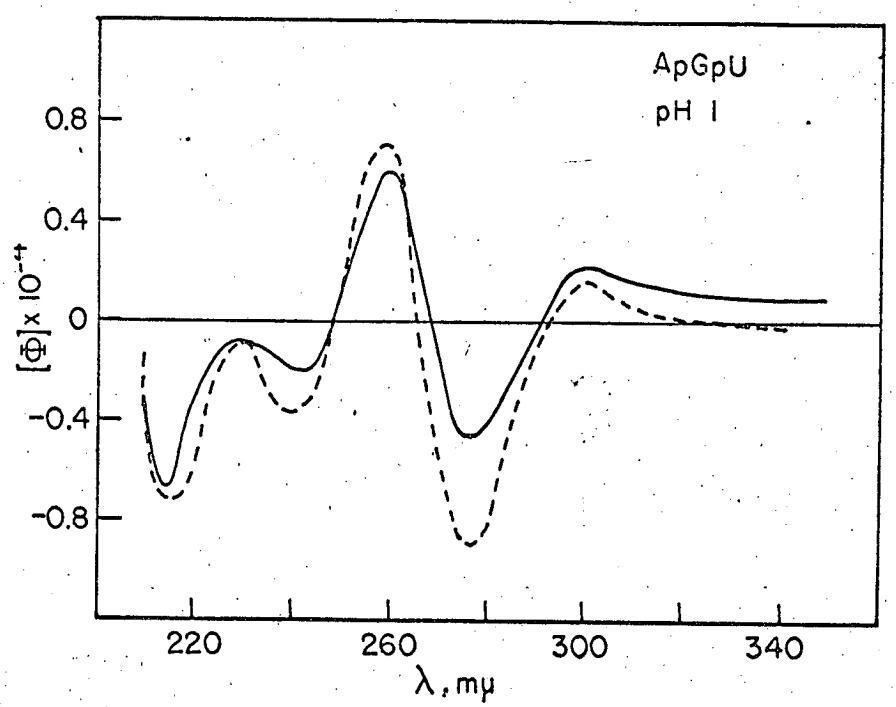


Figure 10. Optical rotatory disperison of ApGpU at pH 1.

——— experimental

– – – – – nearest neighbor calculation

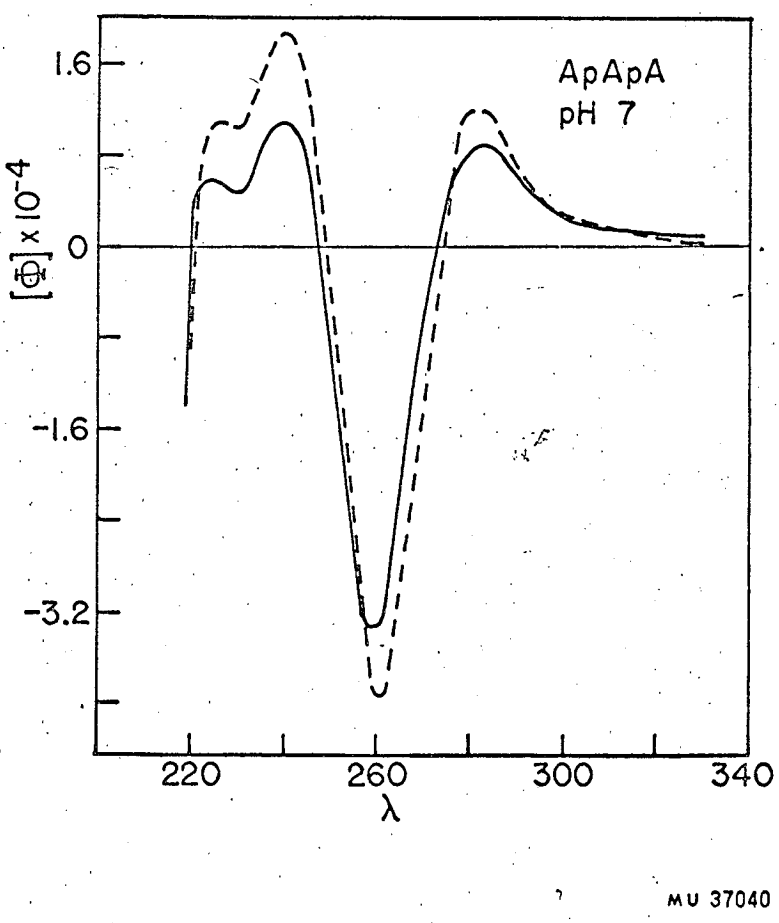


Figure 11. Optical rotatory dispersion of ApApA at pH 7.

——— experimental

– – – – – nearest neighbor calculation

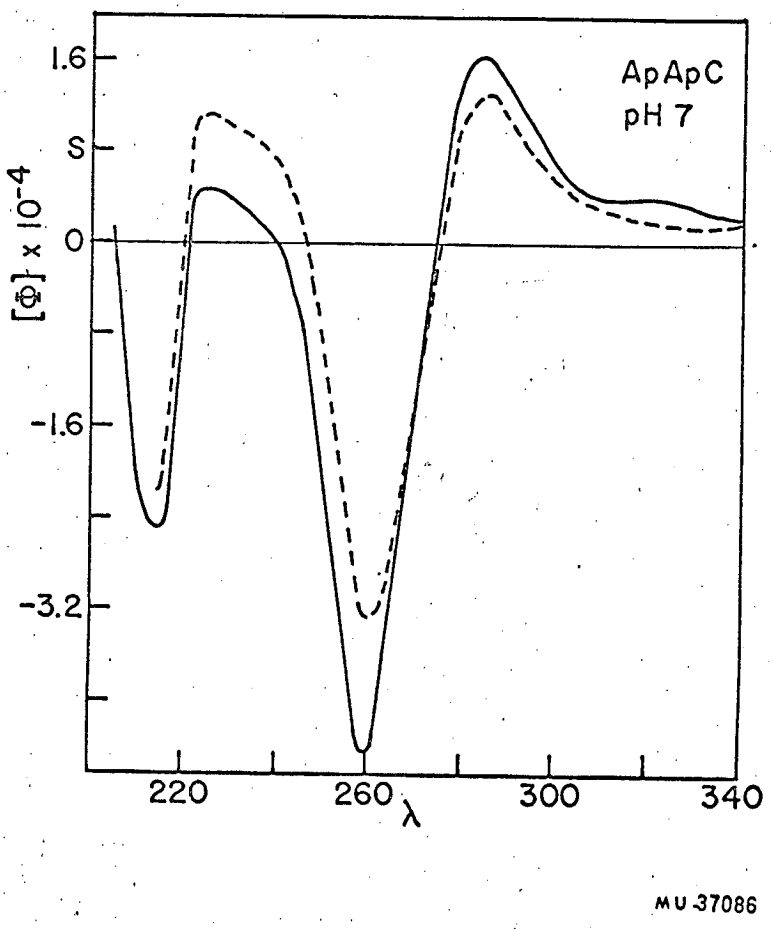


Figure 12. Optical rotatory dispersion of ApApC at pH 7.

——— experimental

– – – – – nearest neighbor calculation

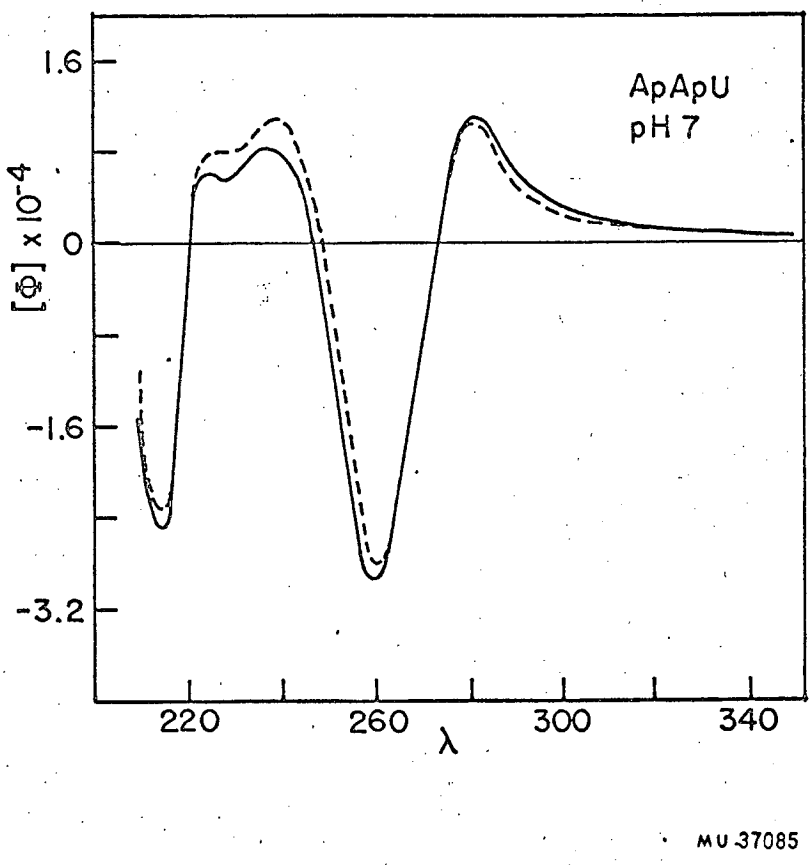


Figure 13. Optical rotatory dispersion of ApApU at pH 7.

——— experimental

– – – – nearest neighbor calculation

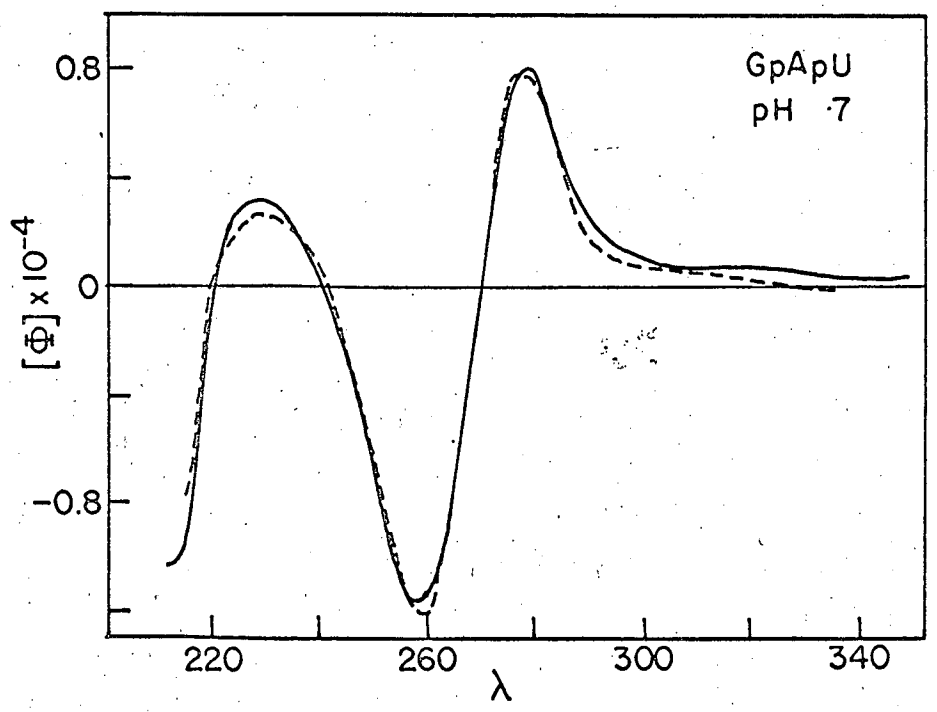


Figure 14. Optical rotatory dispersion of GpApU at pH 7.

——— experimental

– – – – – nearest neighbor calculation

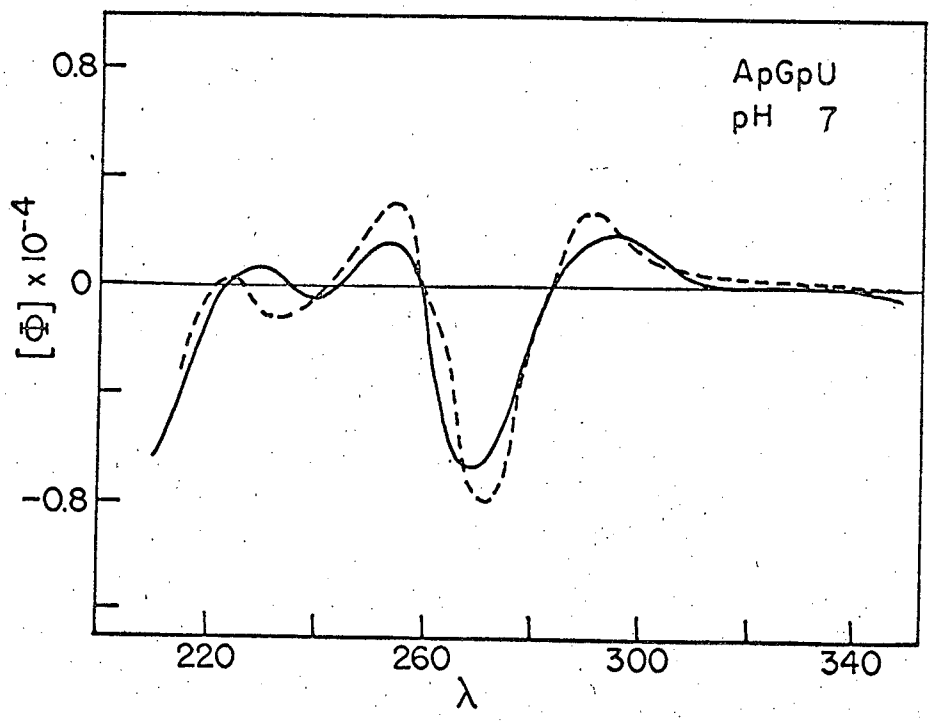


Figure 15. Optical rotatory dispersion of ApGpU at pH 7.

——— experimental

– – – – – nearest neighbor calculation

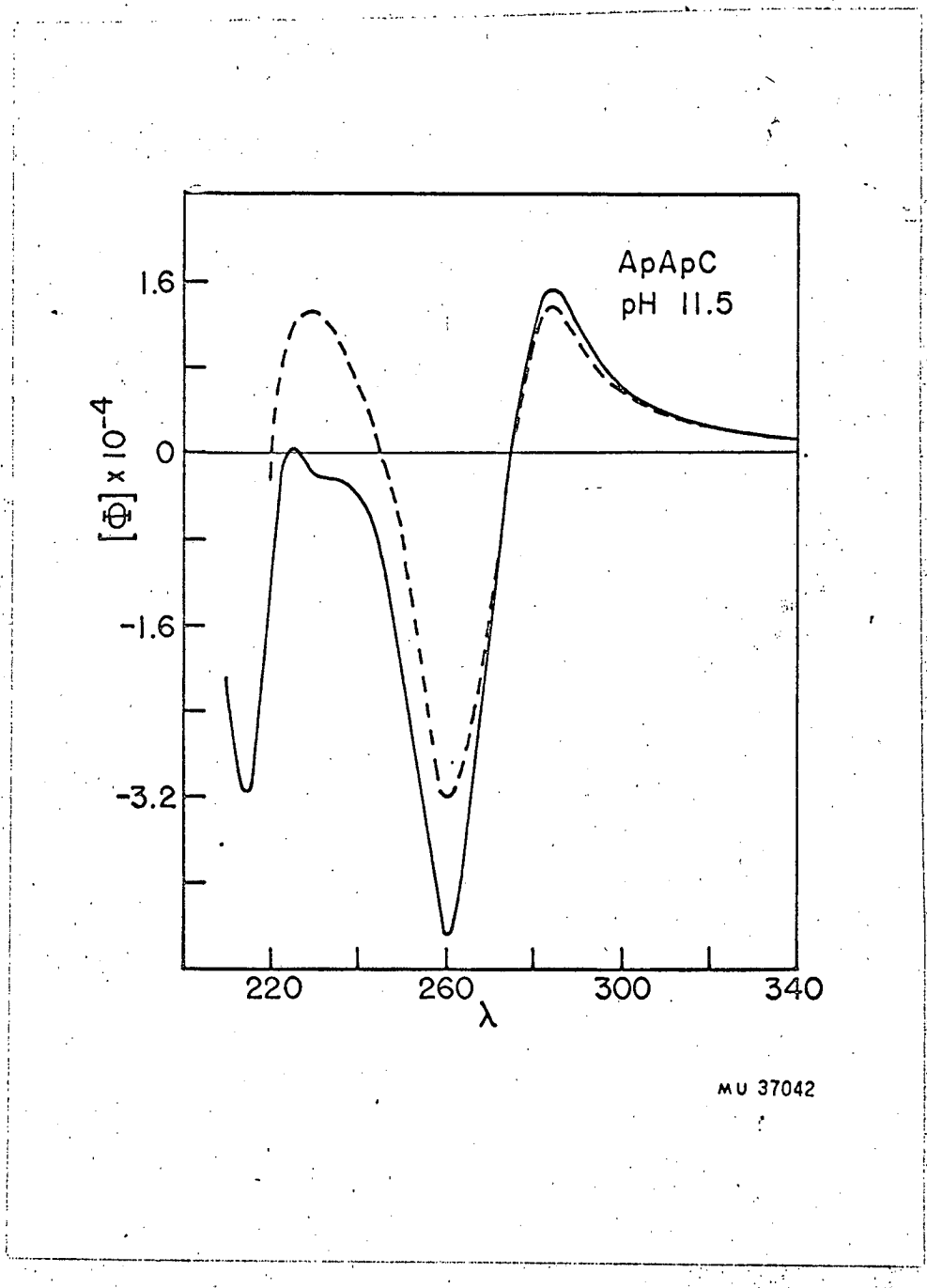


Figure 16. Optical rotatory dispersion of ApApC at pH 11.5.

——— experimental

– – – – – nearest neighbor calculation

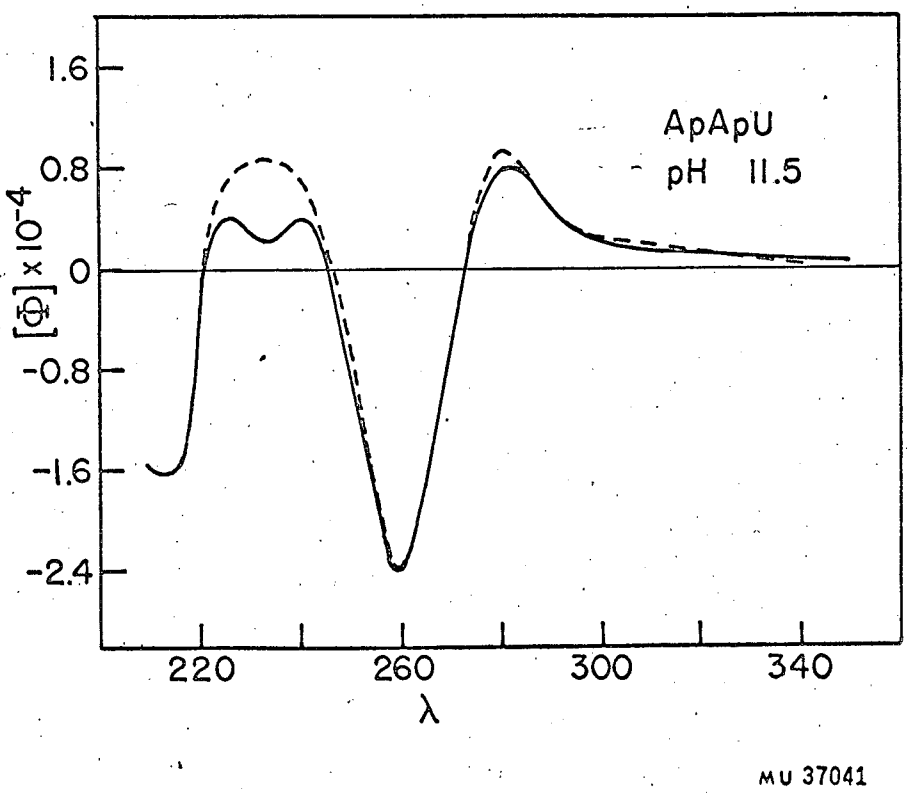
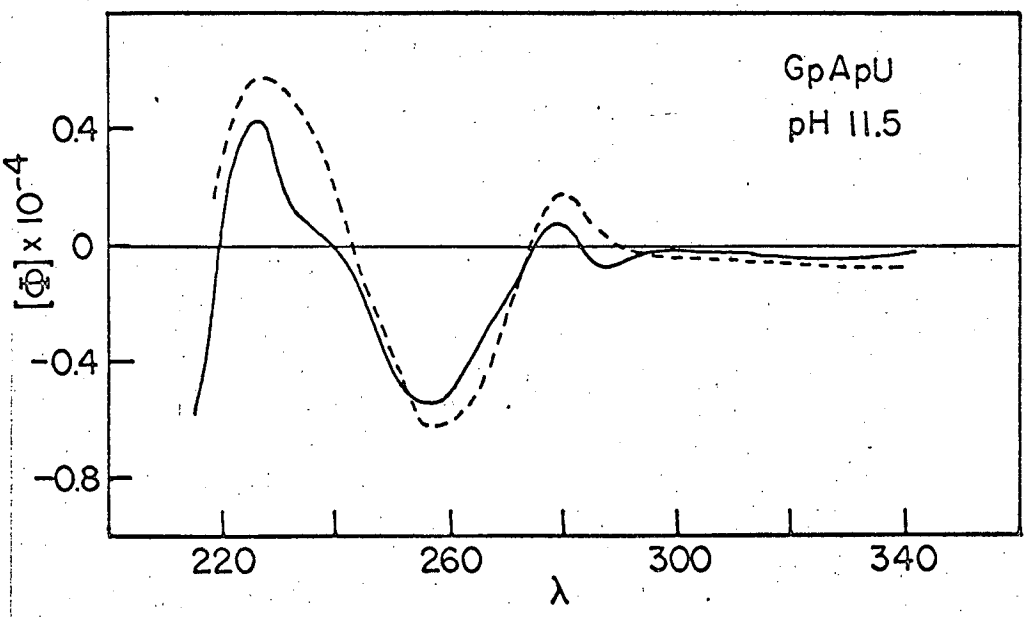


Figure 17. Optical rotatory dispersion of ApApU at pH 11.5.

——— experimental

– – – – – nearest neighbor calculation



MUB-8751

Figure 18. Optical rotatory dispersion of GpApU at pH 11.5

——— experimental

– – – – nearest neighbor calculation

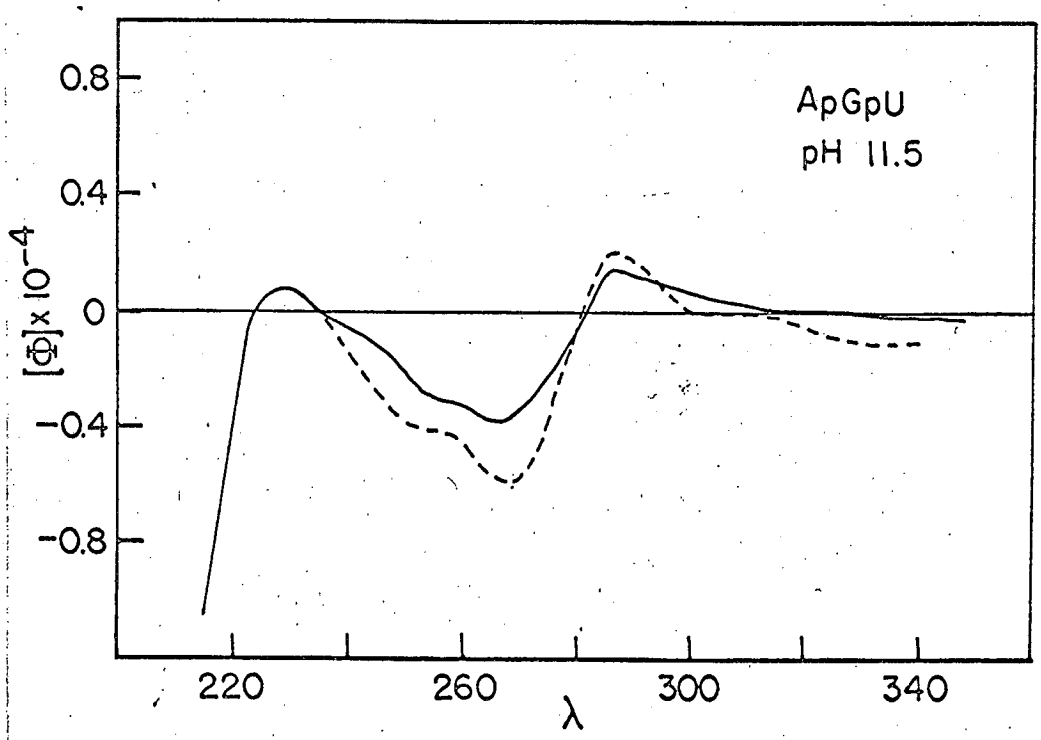


Figure 19. Optical rotatory dispersion of ApGpU at pH 11.5.

——— experimental

– – – – – nearest neighbor calculation

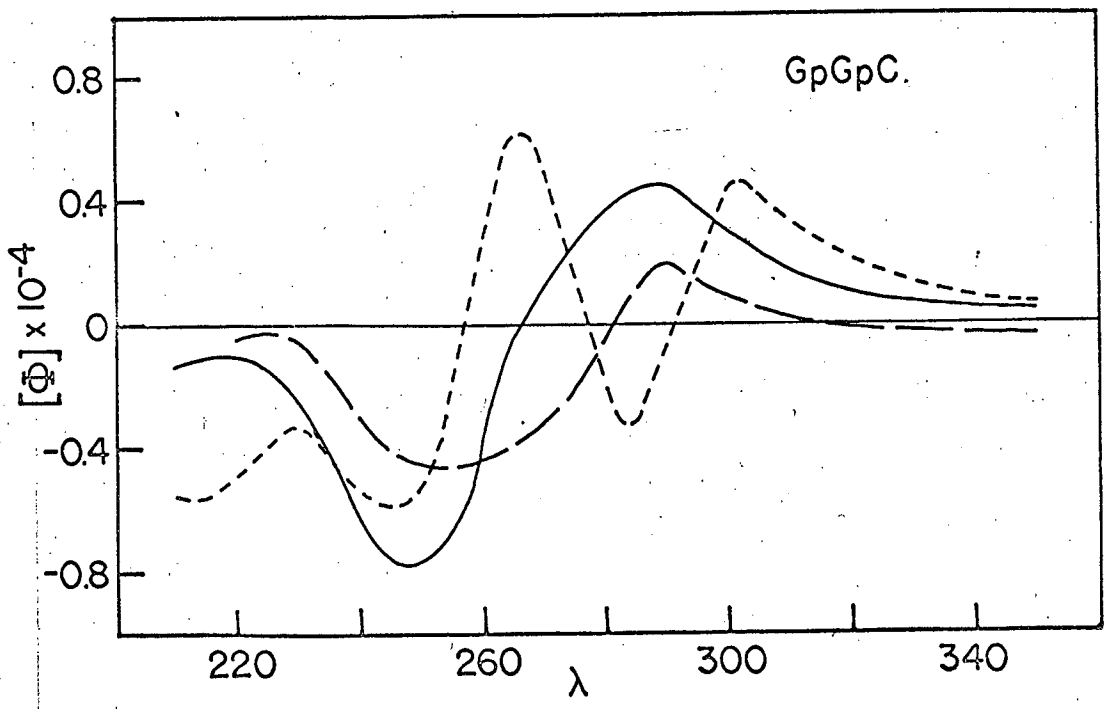


Figure 20. Optical rotatory dispersion of GpGpC.

——— pH 7

— — — pH 11.5

- - - - pH 1

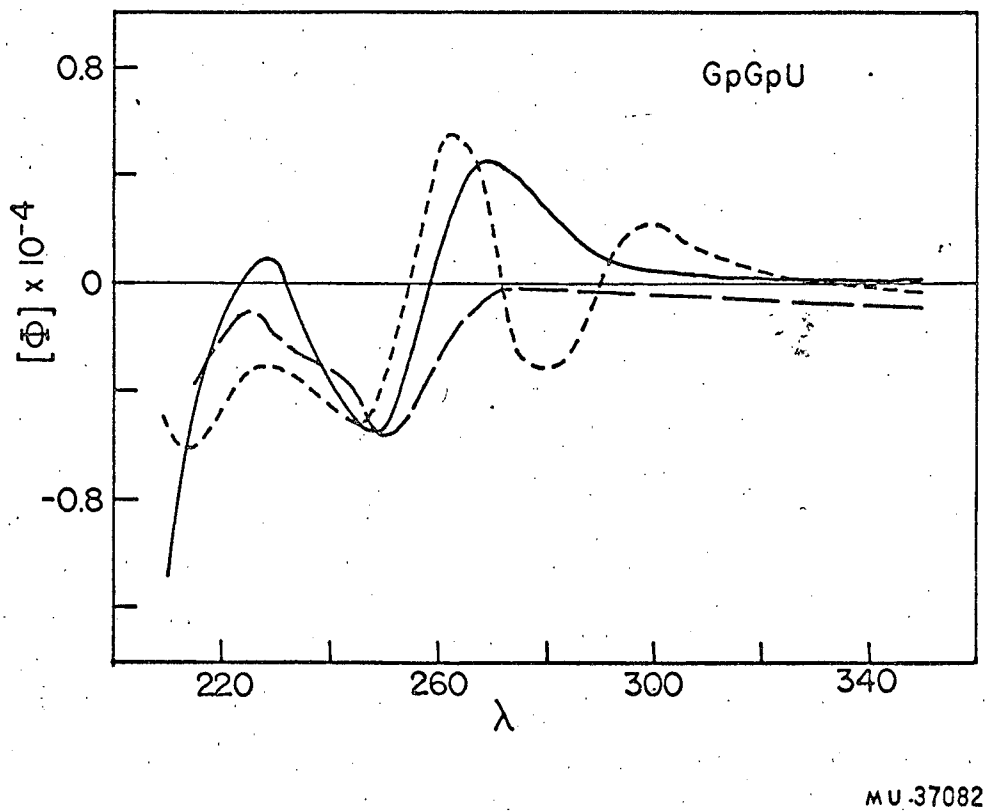


Figure 21. Optical rotatory dispersion of GpGpU.

——— pH 7

— — — pH 11.5

- - - - - pH 1

TABLE I

Optical Rotatory Dispersion of Seven Trinucleoside Diphosphates

Substance	Long Wavelength			Next Wavelength			Additional Wavelengths	
	λ_p	$[\phi_p]\times10^{-4}$	λ_o	λ_t	$[\phi_t]\times10^{-4}$	λ_o	λ_p	$[\phi_p]\times10^{-4}$
					pH 1			
ApApC	300	0.16	290	258	-0.60		242	-0.42
ApApU				257	-0.49	237	232	0.02
GpGpC	302	0.46	291	283	-0.32	277	265	0.62
GpApU	287	0.19	279	251	-0.39		232	-0.14
ApGpU	295	0.19	290	277	-0.46	268	259	0.58
GpGpU	300	0.23	290	280	-0.32	271	263	0.55
					pH 7			
ApApA	283	0.90	272	259	-3.08	247	240	1.08
ApApC	285	1.68	275	260	-4.43	241	225	0.50
ApApU	282	1.11	273	259	-2.92	247	237	0.80
GpGpC	287	0.46	265	248	-0.78		218	-0.10
GpApU	278	0.82	269	258	-1.17	240	230	0.31
ApGpU	295	0.18	283	268	-0.68	259	254	0.16
GpGpU	270	0.45	259	248	-0.56	232	228	0.09
					pH 11.5			
ApApC	285	1.56	275	260	-4.47	226	225	0.04
ApApU	282	0.82	273	259	-2.40	245	240	0.36
GpGpC	290	0.19	282	255	-0.48		226	-0.02
GpApU				255	-0.55	240	226	0.43
ApGpU	285	0.14	281	265	-0.40	235	230	0.08
GpGpU				250	-0.57		225	-0.11

Note: t, p, and o refer to trough, peak, and cross-over, respectively.

studies on oligonucleotides.[248] Thus it seems unlikely that intermolecular interactions are responsible for the ORD of trimers observed under our experimental conditions.

In the case of trimers containing several different bases, the ORD results from the combination of many double Cotton effects. Most of these are due to the interaction of transition moments on a given base with nearby transitions on neighboring bases. Additional contributions to the rotation come from the rotational strengths of the bases themselves due to perturbation by the asymmetry of the ribose phosphate backbone. To a good approximation, this should resemble the monomer rotation. In trimers like ApApA, which contain only one type of base, the double Cotton effect arises from the so-called degenerate interaction of transition moments corresponding to absorption bands which have been split due to exciton interaction.[25] The most striking thing about the ORD of most of the trinucleoside diphosphates at pH 7 is how different it is from the sum of the rotation of the monomers. This shows that the major contribution must come from base-base interactions. Thus we feel that all of these trimers have stacked bases at pH 7.

The ORD of ApApC and ApApU is an order of magnitude weaker at pH 1 than pH 7, and is almost identical with the sum of the molar rotations of the constituent mononucleotides (Figures 7 and 8). Therefore, we think that at pH 1 ApApC and ApApU have an essentially random conformation. Since the bases A and C have a positive charge at this pH, the charge repulsion will counteract the stacking energy. This same effect has been found in the dinucleoside phosphates ApA, ApC, and many others.[252] It is now known, however, at what pH the bases A and C become protonated in a trinucleoside diphosphate.

GpGpC and GpGpU have more complicated ORD curves (Figures 20 and 21). This is to be expected since G has at least two absorption bands in the 260 mμ region. The rotatory dispersion of these two trinucleoside diphosphates at pH 1 is qualitatively very similar and closely resembles that of GpC and GpU at this pH. But the ORD of GpGpU and GpGpC at this pH is very different from the sum of the rotations of their component mononucleotides. This indicates that the major part of the ORD is probably due to base-base interaction, and suggests that these compounds are stacked at pH 1. In contrast, the ORD of GpGpU and GpGpC at pH 7 is roughly a single Cotton effect with a cross-over at 266 mμ for the latter and 259 mμ for the former. This is certainly, however, a greatly oversimplified picture of the ORD of these compounds.

At pH 11.5, the ORD of ApApC and ApApU is almost exactly the same as at pH 7, which indicates that these trinucleoside diphosphates are still stacked at this pH (Figures 16 and 17). The ORD of GpApU and ApGpU at pH 11.5 is very small, and shows a broad trough centered at about 255-260 mμ (Figures 18 and 19). This suggests that these compounds are mostly unstacked at pH 11.5, since the characteristic double Cotton effect, present at pH 7, has disappeared. At pH 11.5, the ORD of GpGpU and GpGpC is almost identical to the sum of the ORD of the respective mononucleotides. Therefore we suggest that the bases of these compounds are not stacked at this pH.

The ORD of GpApU and ApGpU at pH 7 are very different (Figures 14 and 15). The former has a cross-over at 269 mμ and a trough at 258 mμ, while the latter has a trough at 269 mμ and a cross-over at 259 mμ. Thus an unknown mixture of the two of them could easily be quantitatively

analyzed. This shows that the ORD is indeed dependent on the base sequence. Later in this chapter it is shown that it is possible to determine the sequence of a trinucleoside diphosphate of known base composition just by measuring its ORD. At pH 1 the ORD of the two sequence isomers is also very different. That of ApGpU very closely resembles the ORD of GpU at pH 1, while the ORD of GpApU is approximately the sum of the ORD of its mononucleotide components (Figures 9 and 10). Thus GpApU is unstacked at pH 1. The conformation of ApGpU at this pH is probably partially stacked.

In all of the above cases the conformation predicted from a measurement of the ORD of the trinucleoside diphosphates is fairly consistent with the conformation predicted for their component dinucleoside phosphates from the ORD and absorption of the latter.[254]

The maximum difference that occurs in the range of 240 mμ to 350 mμ between the molar rotation per residue of each of the trinucleoside diphosphates studied and the average of the rotations of the mononucleosides it contains is shown in Table I as $|[\phi]_T - [\phi]_M|$. Similar data have been obtained for 16 dinucleoside phosphates.[254] The distribution of the values of this maximum difference in rotation for these dimers and trimers is as follows:

$\lvert[\phi]_T - [\phi]_M\rvert \times 10^{-4}$		0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0	>1.0
Number of compounds	Dimers	7	11	7	10	1	12
	Trimers	3	3	3	4	1	5

Thus the values of the increment in rotation seem to fall into three groups. This distribution suggests that the ORD is arising from three different kinds of conformations. Cantor and Tinoco originally suggested that the

compounds with differences in rotation greater than 1.0×10^4 had stacked conformations. Those with differences less than 0.4×10^4 were probably unstacked, and those with intermediate values were probably partially stacked.[26] These predictions of the conformation of the trinucleoside diphosphates are shown in Table III.

The above treatment is open to question on several grounds. While it is a good first approximation, it ignores the possibility that large rotational bands may cancel, leaving a small net rotation which is not very different from the monomers even though the conformation may be stacked. Thus while the above distribution suggests that GpGpC and GpGpU are unstacked at pH 7, we believe that their small rotation does arise from the cancelling of bands, and thus think that they are actually stacked at this pH.[252]

A second and more powerful objection to the above simple analysis is that it assumes that the rotational strengths of the fully stacked dimers and trimers are all about the same magnitude. This assumption may be very far from the truth, and recent data by Davis suggests that, in fact, dimers containing U may be as stacked as those without U, even though the molar rotation of many of them is very similar to that of the monomers.[37] It is fair to say that, while the presence of a strong ORD different from the monomers is indicative of a stacked conformation, the lack of this rotation does not necessarily mean that the oligomer is unstacked. Once the dimers have been studied at low temperature, and the rotation of the presumably fully stacked conformation measured, this uncertainty can be overcome.

One last objection to the above oversimplification is that we have ignored the possibility of equilibria among the various possible different stacked conformations, which may have very different rotational strengths. The data at present do not seem to warrant corrections to account for these added complications, but they should be kept in mind.

8. Experimental Results--UV Spectra

The absorption spectra of ApGpU and GpApU at 3 pH's are shown in Figures 22 through 27, along with calculated curves, which will be explained later, and curves which were the sum of the monomer absorption. For brevity, only the absorption curves for these two trimers are shown, since the spectra of all eight pancreatic ribonuclease trinucleotides are available in the thesis of Stanley.[222] Relevant spectral data for the trinucleoside diphosphates obtained in our laboratory are summarized in Table II. The absorption spectra of the nucleotides used below were obtained from Warshaw.[251]

From the absorption curves shown in Figures 22 and 23, it can be seen that the absorption curve at pH 1 is slightly broader and less intense in the trinucleosides than in the sum of their component monomers. This is true to a greater extent on the spectra at pH 7, shown in Figures 24 and 25, and the pH 11.5 spectra of Figures 26 and 27. The changes in absorbance at the three pH's seem to be much less spectacular than the corresponding changes in ORD, although we have not made the proper quantitative comparisons. Furthermore, the differences in UV spectra are mostly qualitative, whereas the ORD of the trimers in some cases bears absolutely no resemblance to the monomer rotation. Thus ORD seems to be a much

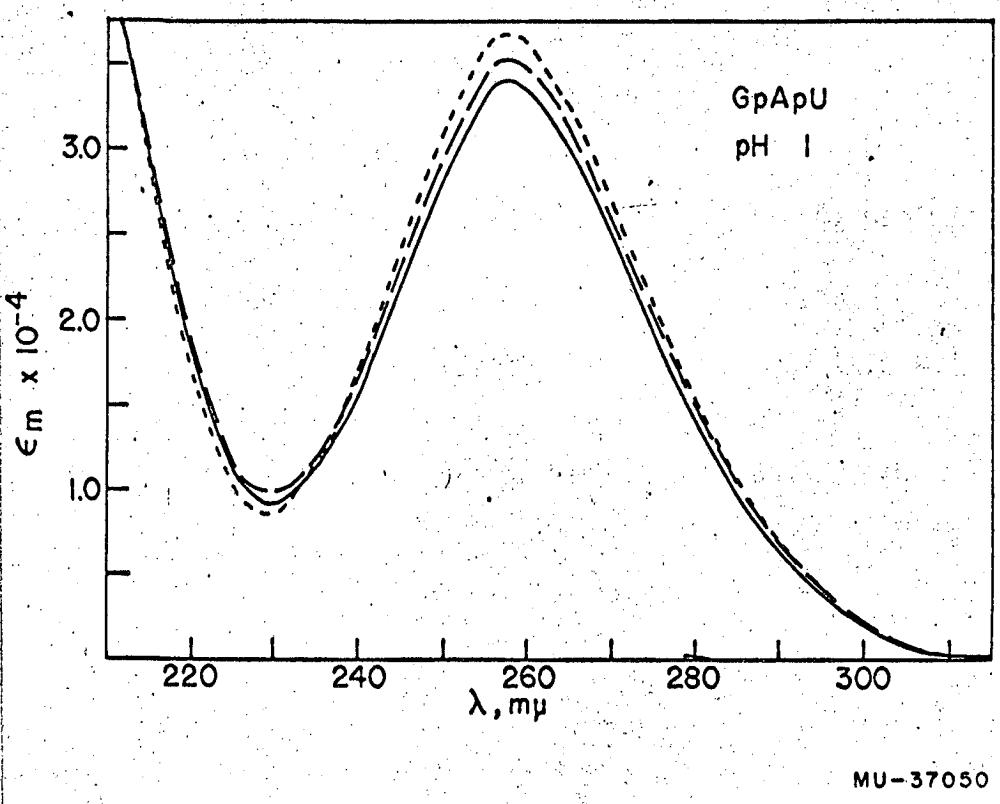


Figure 22. Absorption spectrum of GpApU at pH 1, in units of molar extinction.

——— experimental

— — — nearest neighbor calculations

— · — · — sum of monomers

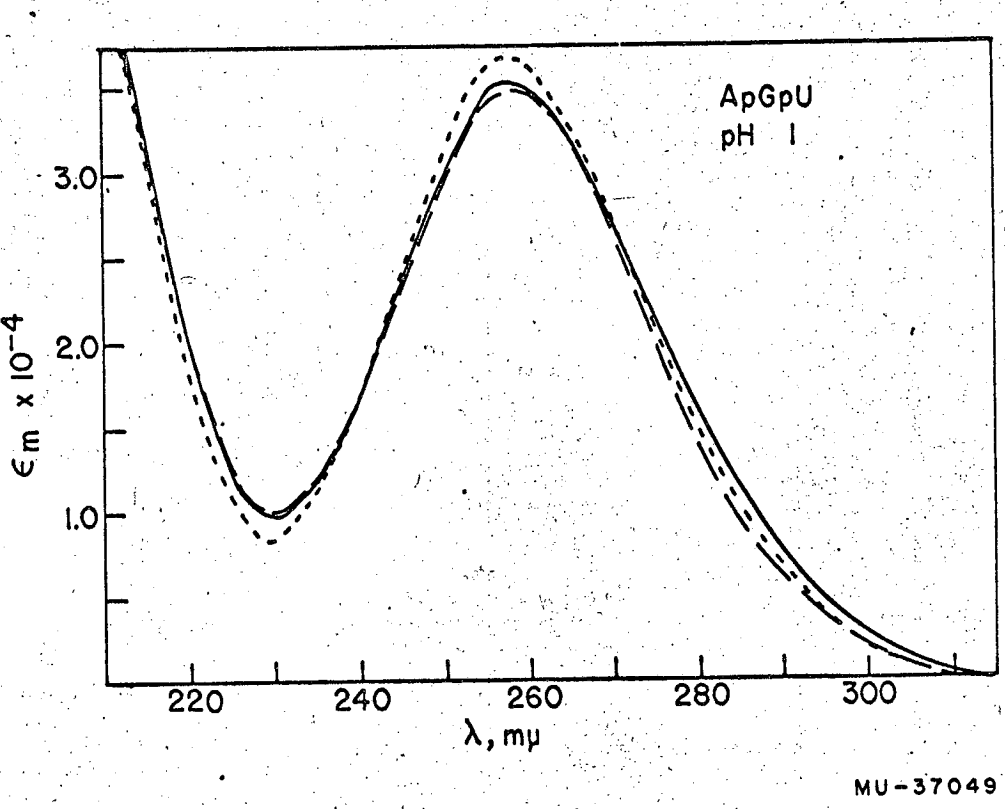


Figure 23. Absorption spectrum of ApGpU at pH 1, in units of molar extinction.

——— experimental

— — — nearest neighbor calculations

- - - - - sum of monomers

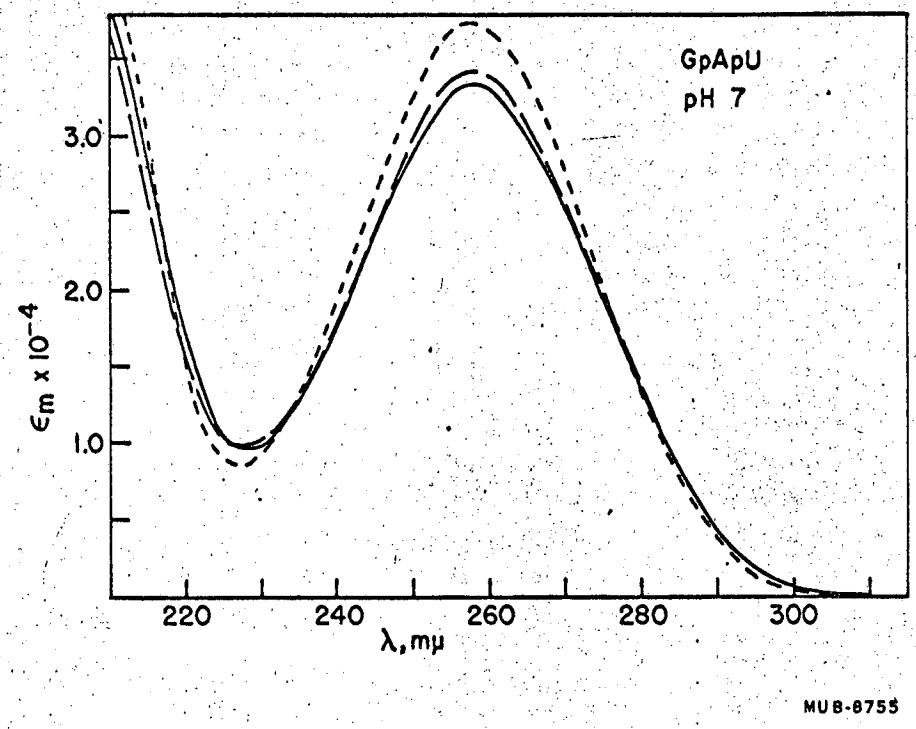


Figure 24. Absorption spectrum of GpApU at pH 7, in units of molar extinction.

——— experimental

— — — nearest neighbor calculations

- - - - - sum of monomers

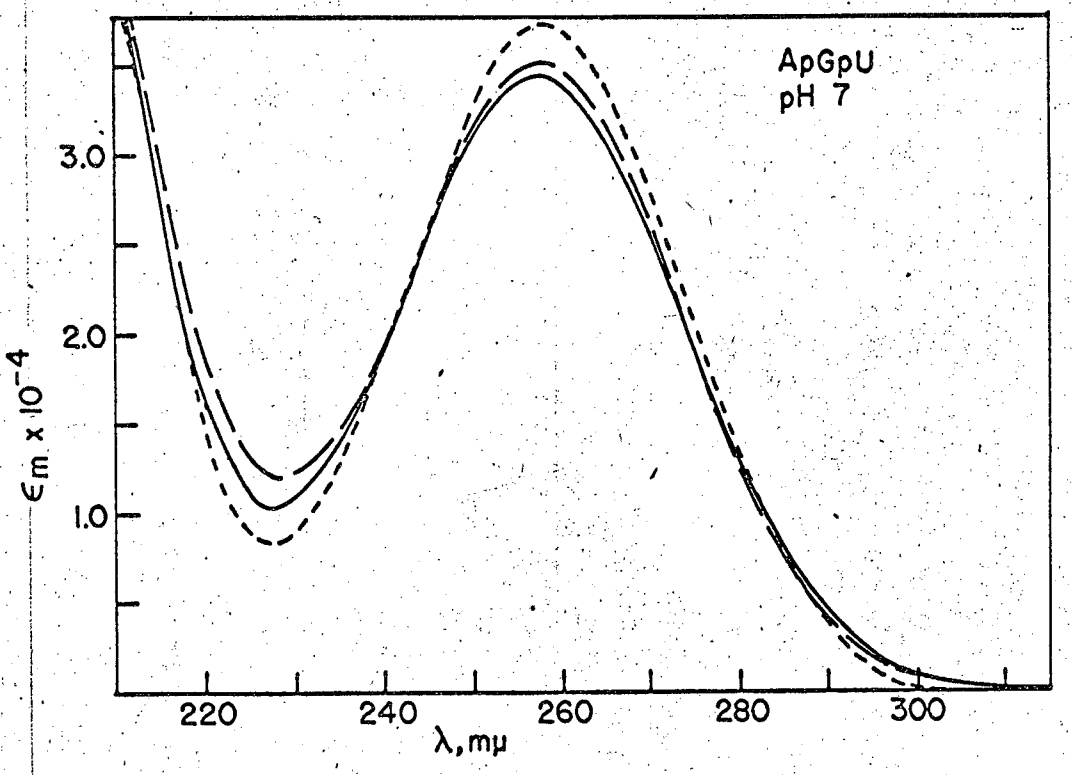


Figure 25. Absorption spectrum of ApGpU at pH 7, in units of molar extinction.

——— experimental

— — — nearest neighbor calculations

- - - - - sum of monomers

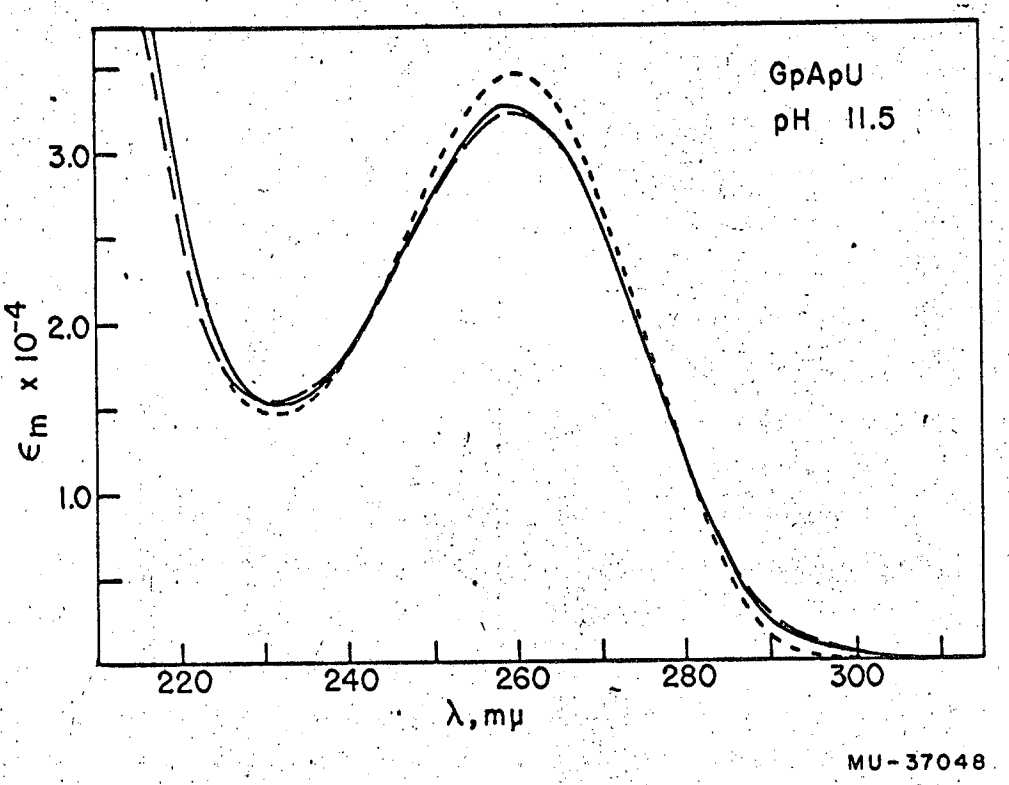


Figure 26. Absorption spectrum of GpApU at pH 11.5, in units of molar extinction.

——— experimental

— — — nearest neighbor calculations

– – – – – sum of monomers

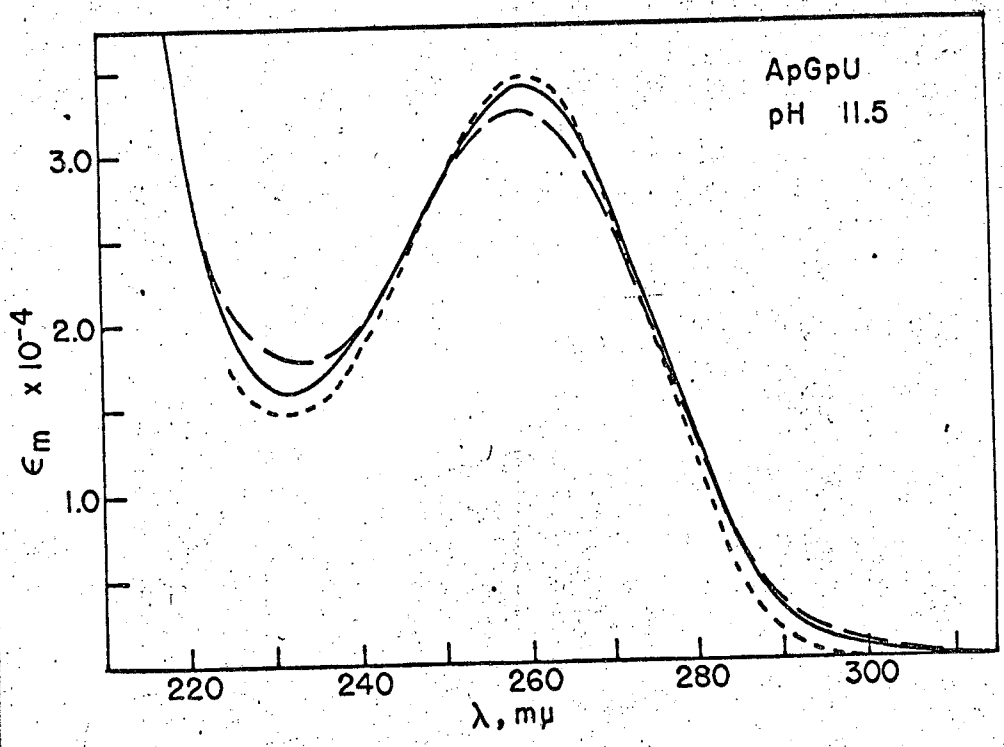


Figure 27. Absorption spectrum of ApGpU at pH 11.5, in units of molar extinction.

——— experimental

— — — nearest neighbor calculations

—·— — sum of monomers

TABLE II

Absorption Spectra of Seven Trinucleoside Diphosphates

Substance	λ_{max}	$\epsilon_{max} \times 10^{-4}$	$\epsilon_{260} \times 10^{-4}$		% Hypochromism	
			Observed	Calc'd	Observed	Calc'd
			pH 1			
ApApC	259	1.13	1.13	1.17	0.1	-0.7
ApApU	257.5	1.26	1.24	1.28	1.0	1.2
GpGpC	260	0.99	0.99		4.2	
GpApU	258.5	1.13	1.12	1.17	7.7	3.3
ApGpU	258	1.17	1.16	1.15	2.9	4.3
GpGpU	258	1.11	1.10		1.4	
			pH 7			
ApApA	258	1.29	1.28	1.30	7.9	8.8
ApApC	259	1.02	1.02	1.09	11.1	9.5
ApApU	258.5	1.14	1.13	1.19	7.3	5.9
GpGpC	255	0.97	0.92		13.0	
GpApU	257.5	1.11	1.10	1.13	8.9	6.2
ApGpU	257	1.15	1.13	1.16	3.9	2.6
GpGpU	255	1.12	1.05		5.2	
			pH 11.5			
ApApC	259	1.04	1.04	1.09	8.2	12.6
ApApU	258	1.11	1.10	1.14	6.8	7.5
GpGpC	266	0.96	0.94		6.0	
GpApU	259.5	1.08	1.08	1.08	2.0	2.1
ApGpU	259	1.13	1.13	1.09	-2.5	-1.0
GpGpU	261.5	1.06	1.06		-5.6	

better method of typing oligonucleotides than absorption spectra. This point is further illustrated by Figures 28 and 29. Here it can be seen that at pH 7, ApGpU and GpApU have entirely different ORD but almost identical absorption spectra.

The hypochromism of the seven trinucleoside diphosphates studied is shown in Table II. It was computed using the equation

$$H = 1 - \frac{f_{ApBpC}}{f_A + f_{pB} + f_{pC}}$$

where f is the oscillator strength. The values of H obtained in this way are more meaningful than just using the decrease in extinction at one wavelength, since they take into account any spectral shifts as well as the changes in the shape of the absorption band. It should be noted that the above equation for the hypochromism is approximate, since it is not obvious which monomer spectra should be correctly compared with the trinucleoside diphosphate. The nucleotide at pH 7, for example, has two negative charges and a very slightly different spectrum from the nucleoside which is uncharged at that pH. Neither of these two is identical with the spectrum of a base surrounded by two secondary phosphates with one charge each like pBp of the above trimer. If the UV spectra of the various species A, Ap, pA, pAp, etc. were very different, we would be in serious trouble unless we could obtain the UV spectrum of a trinucleoside diphosphate in which two of the bases had been removed. Fortunately, the UV spectra of the monomers seems to be very insensitive to the type (if any) of phosphate attached to the sugar. Thus for simplicity we used the monomers shown in the above equation.

To compute H for an isolated absorption band, one would integrate over the entire band to obtain the various oscillator strengths. Since the long wavelength band is not well resolved from the next one for any of the oligonucleosides or mononucleosides we have studied, we decided, as discussed before, to use an arbitrary short wavelength cut-off. By choosing 232.5 mμ for spectra at pH 1 we essentially consider only the longest wavelength absorption band of C, A, and U, and the two long wavelength bands of G. At pH 7 we used a 230 mμ cut-off which therefore includes the second absorption band of C as well as all those included at pH 1. The 232.5 mμ cut-off chosen for the pH 11.5 spectra included all the bands used for calculating the pH 7 oscillator strengths as well as part of the second longest wavelength band of U. These choices are consistent within each pH, but they make comparison of hypochromisms measured at different pH's less meaningful. These choices of cut-offs were chosen to try to include most of the long wavelength bands of all of the trinucleoside diphosphates, dinucleoside phosphates and mononucleosides studied, and thus represent a compromise from the most reasonable cut-off for a given compound.

If we consider the distribution of hypochromism of the trinucleoside diphosphates and the dinucleoside phosphates, the range of hypochromism measured is from -5.6% to +13.0%. The distribution of the values of the hypochromism for the 39 compounds and pH's discussed previously is as follows:

%H		<1,	1-2,	2-3,	3-4,	4-5,	5-6,	6-7,	>7
Number of Compounds	Dimers	16	7	7	0	5	3	4	6
	Trimers	3	2	2	1	1	1	2	7

Thus it can be seen that the values of H seem to fall into two main groups, with a few compounds having hypochromisms between 3% and 6%. Since base stacking would be expected to cause hypochromic spectra relative to the monomer absorption, we feel that these data indicate that oligonucleosides having greater than 6% hypochromism have a conformation in which the bases are stacked. Compounds that have less than 3% hypochromism have unstacked bases, and compounds with H between 3% and 6% are probably partially stacked. This division into three types of conformations is similar to that done before from the ORD data. It suffers from many of the same assumptions, reinforced by great experimental inaccuracy. However, in using hypochromism as a criterion of stacking, we are on firmer ground in some respects. Since almost all of the intense absorption in the bands lumped into the 260 transitions in nucleotide bases are thought to be π-π^*, we expect that contributions from different bands should have more of a tendency to add rather than cancel in the long wavelength part of the spectrum. Thus GpGpC whose rotation is very like the monomers shows the largest hypochromism of any trimer, 13%. Since the bases all have relatively similar polarizabilities,[40] we can expect that the hypochromism of a completely stacked oligomer will depend much less on the particular bases involved than the ORD.

Predictions of the conformations based on the above interpretation of the hypochromism data are shown in Table III. In almost all cases the conformation obtained from the hypochromism agrees with that obtained from the ORD. This lends additional credence to our general picture of the conformation of the oligonucleosides. The only serious exceptions

TABLE III

Conformation of the Trinucleoside Diphosphates

pH 1

	ApApA	ApApC	ApApU	GpApU	ApGpU	GpGpC	GpGpU
Predicted	U	U	U	U	U	S	$\frac{1}{2}$S
ORD	-	U	U	U	$\frac{1}{2}$S	$\frac{1}{2}$S	$\frac{1}{2}$S
Absorbance	-	U	U	S	U	$\frac{1}{2}$S	U

pH 7

	ApApA	ApApC	ApApU	GpApU	ApGpU	GpGpC	GpGpU
Predicted	S	S	$\frac{1}{2}$S	$\frac{1}{2}$S	$\frac{1}{2}$S	S	$\frac{1}{2}$S
ORD	S	S	S	$\frac{1}{2}$S	$\frac{1}{2}$S	U	U
Absorbance	S	S	S	S	$\frac{1}{2}$S	S	$\frac{1}{2}$S

pH 11.5

	ApApA	ApApC	ApApU	GpApU	ApGpU	GpGpC	GpGpU
Predicted	S	S	$\frac{1}{2}$S	U	U	U	U
ORD	-	S	S	U	U	U	U
Absorbance	-	S	S	U	U	S	U

U = unstacked

S = stacked

$\frac{1}{2}$S = partially stacked

are GpGpC at pH 11.5, and GpApU and GpGpU at pH 1. The conformations predicted for the trinucleoside diphosphates using the known conformations of their component dinucleoside phosphates[254] are also shown in Table III. As can be seen, the agreement among the three different methods is generally very good. The excellent agreement at pH 7 gives support to the working hypothesis that neutral A, G and C stack, but U does not. This is reinforced by the hydrodynamic data of Richards, Flessel, and Fresco,[182] which shows that poly U behaves very much like a random coil. It should be stressed that the division of conformations into three classes in Table III is certainly an oversimplification. There is certainly a continuum of conformations depending on the temperature and pH.

If the hypochromicity at 260 mμ is used,

$$H(260) = 1 - \frac{A_{IpJpK}(260)}{A_I(260) + A_{pJ}(260) + A_{pK}(260)}$$

the same qualitative conclusions are obtained in most cases, but since spectral shifts do occur in going from the monomers to the trinucleoside diphosphates, H(260) is not as good a measure of the conformation of these compounds as H. In defining the hypochromicity in the above fashion, we have made the same approximations we discussed in reference to hypochromism.

In Table IV, the hypochromicity at 260 mμ, which we obtained for six trinucleoside diphosphates, is compared with values found by other workers for the same compounds,[219,151] and the values found by W. Stanley, Jr.[222] for the hypochromicity of the corresponding trinucleotides.

TABLE IV

% Hypochromicity at 260 mμ

Compound	pH	This Work	Staehelin*	Stanley**	Michelson***	Calculated
ApApC	1	4.5		2.0		0.1
ApApU	1	5.6		5.8		1.0
GpGpC	1	1.0		4.7		
GpApU	1	7.9		5.2		3.9
ApGpU	1	4.4		3.3		5.2
GpGpU	1	1.8		5.4		
ApGpC	1			1.1		4.6
GpApC	1			3.0		2.8
ApApA	7	16.9				15.1
ApApC	7	19.9	15.5	18.6	18.9	11.1
ApApU	7	16.4	12.1	15.5		12.0
GpGpC	7	9.8	6.5	7.9		
GpApU	7	10.6	7.7	9.5	10.1†	8.3
ApGpU	7	7.9	8.2	6.7		5.7
GpGpU	7	4.4	4.2	6.0		
ApGpC	7		10.3	9.6		9.8
GpApC	7		10.9	10.8		10.3
ApApC	11.5	17.6		18.4		14.2
ApApU	11.5	14.8		11.9		11.4
GpGpC	11.5	5.7		2.3		
GpApU	11.5	6.6		4.9		6.6
ApGpU	11.5	2.3		2.6		5.8
GpGpU	11.5	-3.9		-5.6		
ApGpC	11.5			6.1		5.8
GpApC	11.5			8.4		10.7

(Cont.)

Table IV (Cont.)

*Reference 219.

**Reference 222. These compounds contained a 3' terminal phosphate.

***Reference 151.

†Contains mixed (2'-5') and (3'-5') phosphates.

The calculated values will be discussed later. It can be seen, by comparing the work of Stanley with the other results, that the addition of a 3' terminal phosphate does not substantially change the hypochromicity of a trinucleoside diphosphate. Measurements on ApApAp and ApApA[38] show that these compounds have almost the same ORD, which further suggests that the effect of a terminal phosphate on the optical properties of trinucleoside diphosphates is small. But data obtained in our laboratory by S. Davis show that a dinucleoside phosphate can have substantially different ORD than the corresponding dinucleotide.[38]

The finding that the terminal phosphate may not affect the conformation of a trinucleotide is interesting in the light of the observation of Nirenberg and Leder that the presence of a terminal phosphate, either 5' or 3', can cause marked differences in the messenger activity of trinucleotides.[163] Perhaps, as they suggest, this difference is due to a special role for the terminal phosphate in binding the messenger to the ribosome, and not the result of a change in the conformation of the messenger.

9. Nearest Neighbor Calculations--Principles

Now that experimental data are available for several of the optical properties of trinucleoside diphosphates and dinucleoside phosphates, we shall see whether it is possible to explain the properties of the former in terms of the properties of the latter. If we are successful, we shall then try to extend the argument to polynucleotides.

In an attempt to explain the ORD and hypochromism of the trinucleoside diphosphates, we have developed a simple scheme for calculating the optical properties of polynucleotides from the optical properties of their constituent dinucleoside phosphates. There are two assumptions implicit in the following calculations. First, we assume that the optical properties can be accounted for completely by nearest neighbor interactions between the bases. The second assumption is that the relative conformation of any two bases is the same in the dinucleoside phosphate as it is if the two bases were nearest neighbors in a longer oligonucleotide.

For any dinucleoside phosphate we can write that

$$[\phi'_{IJ}(\lambda)] = [\phi'_{I}(\lambda)] + [\phi'_{J}(\lambda)] + N_{IJ}(\lambda) \qquad (1)$$

where $[\phi'_{IJ}(\lambda)]$ is the molar rotation of the dinucleoside phosphate IpJ, $[\phi'_{J}(\lambda)]$ is the molar rotation of the mononucleoside J or the mononucleotide pJ, and $N_{IJ}(\lambda)$, defined by this equation, is the nearest neighbor interaction contribution to the ORD. In principle, $[\phi'_{J}(\lambda)]$ should be the molar rotation of the dinucleoside phosphate with one base removed; ribose(3'→5')pJ, and $[\phi'_{I}(\lambda)]$ would be the rotation of the corresponding

compound, Ip(3'→5')ribose. We have not yet been able to measure the optical properties of any of these compounds, but we expect that they would have properties very similar to the mononucleotides or mononucleosides, since the perturbing effect of the second ribose will be small because of its relatively large distance from the base, and the absence of a direct covalent link. We are not sure whether a mononucleotide or mononucleoside is the better model, since the former has a primary phosphate which has two negative charges at pH 7 instead of only the one charge present on the secondary phosphate of the dinucleoside phosphate. But the mononucleoside has no phosphate at all. In practice, we usually employ the average of mononucleoside and mononucleotide properties for the above equation. In general, the molar rotations of the mononucleoside and the mononucleoside phosphates are very similar,[127] and small compared to the molar rotation of the dinucleoside phosphate.[252] Thus it does not matter which one is chosen for the above equation. This separation of the molar rotation into three contributions may or may not have any physical significance, but mathematically it is always possible to do this.

For any trinucleoside diphosphate we can write that

$$[\phi'_{IJK}(\lambda)] = [\phi'_{I}(\lambda)] + [\phi'_{J}(\lambda)] + [\phi'_{K}(\lambda)] + M_{IJ}(\lambda) + M_{JK}(\lambda) + M_{IK}(\lambda) \qquad (2)$$

where $[\phi'_{IJK}(\lambda)]$ is the molar rotation of the trinucleoside diphosphate, and $M_{IJ}(\lambda)$ and $M_{JK}(\lambda)$ are the nearest neighbor interactions of those bases as they exist in trinucleoside diphosphates. $M_{IK}(\lambda)$ is the contribution due to next-nearest neighbor interaction. Our first assumption is that $M_{IK}(\lambda) = 0$; the effects of next-nearest neighbor interactions

can be ignored. Next, if the bases in the trinucleoside diphosphates and the dinucleoside phosphates have the same relative conformation, we would expect that the nearest neighbor interaction contributions to the ORD would be the same in both cases. Thus,

$$M_{IJ}(\lambda) = N_{IJ}(\lambda) \quad \text{and} \quad M_{JK}(\lambda) = N_{JK}(\lambda) \ , \tag{3}$$

and using equation (1) for the dinucleoside phosphates IpJ and JpK, we can write equation (2) as

$$[\phi'_{IJK}(\lambda)] = [\phi'_{IJ}(\lambda)] + [\phi'_{JK}(\lambda)] - [\phi'_{J}(\lambda)] \ . \tag{4}$$

If we transform to the units of molar rotation per residue, equation (4) becomes

$$[\phi_{IJK}(\lambda)] = \frac{2[\phi_{IJ}(\lambda)] + 2[\phi_{JK}(\lambda)] - [\phi_{J}(\lambda)]}{3} \ . \tag{5}$$

An exactly analogous calculation for the oscillator strengths or the extinction coefficients leads to the results that

$$f_{IJK} = \frac{2f_{IJ} + 2f_{JK} - f_{J}}{3} \tag{6}$$

and

$$\epsilon_{IJK}(\lambda) = \frac{2\epsilon_{IJ}(\lambda) + 2\epsilon_{JK}(\lambda) - \epsilon_{J}(\lambda)}{3} \tag{7}$$

where ϵ_{IJK} and $f_{IJK}(\lambda)$ are, respectively, the extinction coefficient and oscillator strength per residue of the trinucleoside diphosphate IpJpK. The other terms are defined analogous to the various molar rotations per residue.

By using equation (6) and the definition of the hypochromism, we can derive a useful expression for the hypochromism of trinucleoside

diphosphates in terms of the optical properties of the dinucleoside phosphates and mononulcoosides.

$$H_{IJK} = \left(\frac{f_I+f_J}{f_I+f_J+f_K}\right) H_{IJ} + \left(\frac{f_J+f_K}{f_I+f_J+f_K}\right) H_{JK} . \qquad (8)$$

For the hypochromicity at a given wavelength, the same formal expression is used, but the oscillator strengths per residue and hypochromism are replaced by the extinction coefficients per residue and hypochromicities at that wavelength.

$$H_{IJK}(\lambda) = \left(\frac{\varepsilon_I(\lambda)+\varepsilon_J(\lambda)}{\varepsilon_I(\lambda)+\varepsilon_J(\lambda)+\varepsilon_K(\lambda)}\right) H_{IJ}(\lambda) + \left(\frac{\varepsilon_J(\lambda)+\varepsilon_K(\lambda)}{\varepsilon_I(\lambda)+\varepsilon_J(\lambda)+\varepsilon_K(\lambda)}\right) H_{JK} \qquad (9)$$

In practice, it is easier to calculate the hypochromicity or hypochromism by using the equation

$$H = 1 - f_{IJK} / (f_I+f_J+f_K) \qquad (10)$$

$$H(\lambda) = 1 - \varepsilon_{IJK}(\lambda) / (\varepsilon_I(\lambda)+\varepsilon_J(\lambda)+\varepsilon_K(\lambda)) \qquad (11)$$

where the trimer property is calculated from the appropriate equation.

If end effects are ignored, these results can easily be extended to larger oligomers and polymers.

$$[\Phi_{RNA}(\lambda)] = \sum_{I=1}^{4}\sum_{J=1}^{4} 2x_{IJ}[\Phi_{IJ}(\lambda)] - \sum_{I=1}^{4} x_I[\Phi_I(\lambda)] \qquad (12)$$

where x_{IJ} is the mole fraction of dinucleoside phosphate IpJ, and x_I is the mole fraction of nucleoside I. If the sequence of the polymer is random then the molar rotation per residue becomes

$$[\phi_{RNA}(\lambda)] = \sum_{I=1}^{4} x_I \left\{ 2\left(\sum_{J=1}^{4} x_J [\phi_{IJ}(\lambda)] \right) - [\phi_I(\lambda)] \right\}. \quad (13)$$

Analagous expressions hold for the other optical properties.

10. Nearest Neighbor Calculations--Results

Using equation (5) we have calculated the ORD of ApApA, ApApC, ApApU, GpApU, and ApGpU at various pH's. The dimer data needed for these calculations were obtained from Warshaw.[252] The monomer data are the average of the mononucleosides[38,264] and mononucleotides.[252] The original calculations were done by hand, but more recently we have used the computer program discussed in Appendix 2. The calculated curves are compared with experiments in Figures 7 through 19. We have not yet been able to perform these calculations for GpGpC and GpGpU, since we do not know the molar rotation of GpG. The agreement between experimental and calculated curves is excellent. This suggests that the major contribution to the ORD of the trinucleoside diphosphates comes from the rotation of their dinucleoside phosphates. It also is a strong argument for similar conformations of bases in dinucleoside phosphates and trinucleoside diphosphates. Thus it is possible, by very simple calculations, to predict the ORD of oligonucleotides from their constituent dinucleoside phosphates.

From a comparison of the experimental and calculated curves for ApApA, ApApC, and ApApU it can be seen that there is a consistent discrepancy in the predicted height of the peak in the reagion of 230 mμ at pH 7. In each case, for a trinucleoside diphosphate containing ApA the nearest neighbor calculations predict a greater molar rotation than is actually observed in this region, although the details of the shape of the cruve are quite accurately predicted. This error in magnitude is also present for the ORD of ApApC and ApApU at pH 11.5. It may possibly be due to the influence of the lower wavelength Cotton effect centered at about 215 mμ in ApA at pH 7 or 11.5, since the error is much more pronounced at the short wavelength side of the 260 mμ Cotton effect than at longer wavelengths. In general, of course, one cannot expect the ORD to be solely due to nearest neighbor interactions. For example, the shape of the ORD of poly A cannot be calculated exactly from ApA since its long wavelength trough is at 256 mμ at pH 7, while the corresponding trough of the dinucleoside phosphate is at 260 mμ at this pH. Evidence of this shift is already present in ApApA, where the trough is at about 258.5 mμ.

From a comparison of the calculated and experimental ORD curves for GpApU and ApGpU it can be seen that, not only is ORD sequence dependent, but also that the base sequence can be determined from an ORD measurement. In principle, this can be done by measuring the base composition using any of the standard methods, calculating the ORD for each of the possible sequences consistent with the base composition, and comparing the experimental curve with the calculated ones. The accuracy to which this can be done will depend on the differences among the ORD of the 16

dinucleoside phosphates and the precision of the calculations. Thus far the ORD of 15 of the 16 possible dinucleoside phosphates has been measured in our laboratory, and each of them has an ORD curve significantly different from any of the others.[252]

If all of the 16 dinucleoside phosphates have sufficiently different ORD, then by measuring the ORD of an oligonucleotide of arbitrary length and known base composition one could determine the number of times each of the 16 nearest neighbor base pairs appeared. But this is not enough information to be able to assure the determination of the base sequence. The fraction of all of the possible oligomers of a given chain length whose sequence would be completely specified by a knowledge of the frequency of all nearest neighbor pairs and the base composition is as follows (see Appendix 4):

Chain length	3	4	5	6
Fraction with determinable sequence	64/64	232/256	796/1024	2176/4096

Much information could still be obtained about the sequence of those oligomers whose sequence is not completely determined by nearest neighbor data. Further discussion of the usefulness of ORD for determining sequence is found in Appendix 4.

While ORD is certainly not the most infallible method for determining the sequence of oligonucleotides, its simplicity and the fact that it does not degrade the molecule being studied, certainly warrant its strong consideration as an adjunct to the methods already in use. It should be most useful for trinucleotides and tetranucleotides, since differences

among the ORD of longer sequence isomers may become vanishingly small. Another great advantage of ORD is that 0.7 ml of a 0.7 absorbance solution is all that is needed to make the measurement.

The absorption spectra of ApGpU and GpApU have been calculated from equation (7) and are shown in Figures 22 through 27 along with the experimental curves and the sum of the monomer absorption. In four of the cases the agreement between experimental and calculated curves is excellent. This is very pleasing, since a small error in extinction coefficient could easily have erased most of the differences between monomer and trimer absorption spectra. In two cases, GpApU at pH 1 . and ApGpU at pH 11.5, the agreement is less than desirable. The GpApU data could easily be explained if there was an error (experimental or calculated) in the extinction coefficient. The ApGpU data at pH 11.5 is a little more puzzling since there is excellent agreement between calculated and experimental curves at long wavelength, even though the monomer spectrum is very different from either. For brevity, the other absorption spectra are not shown, but the extent of the agreement is about the same. Results of calculations of spectral ratios are discussed later.

The hypochromism of ApApC, ApApU, ApApA, ApGpU and GpApU has been calculated from equations (6) and (10) [or equation (8)], and the results are shown in Table II. The calculations are in pretty good agreement with experiment. It should be noted that the experimental hypochromism is at most accurate to $\pm$ 1% since it is a small difference between two large numbers. Except in the case of ApGpU at pH 1 and 7, the estimates of the conformation made from the experimental hypochromism are consis-

tent with those that would be predicted from the calculated values. Again, since GpG has not yet been measured in our laboratory, we are unable to calculate the hypochromism of GpGpC and CpGpU. The calculated hypochromism in general agrees better with the experimental values than the calculated hypochromicity at 260 mμ does with the corresponding experimental data shown in Table IV. This is to be expected since spectral shifts do occur in going from the dinucleoside phosphates to the trinucleoside diphosphates, and these would affect the integrated measurement much less than the change in extinction coefficient at a single wavelength. Extinction coefficients calculated from equation (7) are shown in Table II.

The agreement between the calculated hypochromicity shown in Table IV with the experimental values obtained in various labs seems to show no preference for any researcher. This is encouraging since it suggests that the agreement between our experiments and calculations is not due to any experimental idiosyncracies.

Thus we have shown that the ORD and absorption spectra in solution of seven trinucleoside diphosphates can be fairly well accounted for in terms of the corresponding dinucleoside phosphate properties. This indicates that our assumptions may be correct. Thus, we are fairly certain that the conformation of trinucleoside diphosphates in solution is very similar to the conformation of their component dinucleoside phosphates. Furthermore, it shows that, at least for small oligomers, nearest neighbor interactions are sufficient to account for most of the observed optical properties. Whether or not our approach is generally valid awaits the study of many more of these compounds. From these preliminary results,

though, we think that one will be able to learn much about the sequence and conformation of larger oligonucleotides and RNA's from a study of their ORD and absorbance in solution. Thus we have been encouraged to try to extend these results to the optical properties of polymers. This will be discussed in Chapter IV.

11. Predicted ORD of 64 Trimers

Encouraged by our success in calculating the ORD and other optical properties of the seven trinucleoside diphosphates we had studied experimentally, we decided to calculate the optical properties of all 64 trimers. This compilation of data can be useful in several ways. First of all, we hope it will provide an inducement for experimenters who have synthesized many of the trinucleoside diphosphates to measure their ORD. This will greatly facilitate the testing of the assumptions which underlie our nearest neighbor calculations. At the moment we have no intention of synthesizing all 64 trimers, but almost all of them are available in one laboratory or another, albeit in small quantities. The summary of calculated ORD data for trimers at pH 7, shown in Table V, makes it very apparent that ORD is a far better method of identifying and characterizing trinucleoside diphosphates than the corresponding absorption spectra. Note, also, the experimental comparisons shown in Figures 28 and 29. Yet until now only the latter very insensitive method has been used in identifying trimers. We have only calculated data at pH 7, since this is by far the most useful pH for workers interested in the biological properties of these compounds. While the greatest differences in absorption spectra of mixtures of monomers apparently occur at pH 12,[124] we feel that this will not be true for the ORD of oligomers.

TABLE V

Calculated Optical Rotatory Dispersion of 64 Trinucleoside Diphosphates

Molar Rotation x 10^{-4}

Compound	Long Wavelength			Second Wavelength			Third Wavelength			Fourth Wavelength		
	λ_p	$[\phi_p]$	λ_o	λ_t	$[\phi_t]$	λ_o	$\lambda_{(p \text{ or } t)}$	$[\phi]$	λ_o	$\lambda_{(p \text{ or } t)}$	$[\phi]$	λ_o
ApApA	282.5	1.27	274	260	-3.89	249	240	1.86		225(p)	1.01	221.5
ApApU	280	1.03	272.5	260	-2.78	248	238	1.07		227.5(p)	.77	221
ApApC	284	1.31	275	261	-3.25	246	236	.93		226.5(p)	1.13	220
ApApG	285	.53	277.7	262	-2.22	250.5	240	1.16		225(p)	.61	221
ApUpA	277.5	.41	270.7	261	-.93	247	230	.31	219			
ApUpU	281.5	.72	272.5	261.5	-1.53	241.5	228	.41	219.5			
ApUpC	286	.73	273.7	259.5	-1.42	233.5	230	.05	226			
ApUpG	285	.20	276.2	262.5	-.83	251	230	.33	219			
ApCpA	287.5	.98	277	263	-2.21	237	226.5	1.09	219			
ApCpU	287.5	1.47	277	263	-2.42	235.5	227.5	.73	218			
ApCpC	291.5	1.71	279.5	265	-2.31	232.5	227	.59	219.5			
ApCpG	292	.83	280	264.5	-1.80	239.5	227.5	.75	218.5			
ApGpA	287.5	.11	279.2	267.5	-.78	248	240	.16	232.5			
ApGpU	290	.29	283	270	-.78	259	254	.30	241	233(t)	-.15	219
ApGpC	292.5	.58	280.5	267.5	-.86		247.5(p)	-.08	227	218(p)	.40	
ApGpG	290	.12	285	270	-.54		255(p)	-.09	242.5	230(p)	.11	222.5
UpApA	282.5	.84	273.5	260	-2.44	247.7	238(p)	.80		225(p)	.49	221.5
UpApU	278.5	.65	271	260	-1.32	239	227.5	.26	218			
UpApC	285	.90	275.5	262	-1.42	236	227	.61	217.5			
UpApG	290	.25	281.5	267.5	-.94	247	237.5(p)	.09		222(p)	.11	218
UpUpA	282.5	.53	274	260	-1.19	236	227.5	.11	222	217.5(t)	-.27	
UpUpU	284	.93	274.5	260	-1.71	236	227.5	.32	221.5	217.5(t)	-.44	
UpUpC	287.5	1.05	275.5	260	-1.58		227.5(p)	-.15		217.5(t)	-.57	
UpUpG	287	.54	277	263.5	-1.03	237	227.5	.13	222			
UpCpA	287.5	.96	276.7	262.5	-1.68	231.5	227.5	.19	222	216.5(t)	-.70	

TABLE V (Cont.)

Compound	Long Wavelength			Second Wavelength			Third Wavelength			Fourth Wavelength		
	λ_p	$[\phi_p]$	λ_o	λ_t	$[\phi_t]$	λ_o	λ(p or t)	$[\phi]$	λ_o	λ(p or t)	$[\phi]$	λ_o
UpCpU	288	1.45	277	260	-1.94		227.5(p)	-.14		217(t)	-.43	
UpCpC	291	1.68	280	265	-1.80		226.5(p)	-.30		219(t)	-.58	
UpCpG	292	.79	280.5	262.5	-1.24		227.5(p)	-.14		217(t)	-.42	
UpGpA	285	.26	274	260	-.56		230(p)	-.11				
UpGpU	290	.38	279.5	271.5	-.41	257	255(p)	-.02	253.5	240(t)	-.36	
UpGpC	292	.67	277	260	-.61		250(p)	-.43		242.5(t)	-.47	225
UpGpG	287.5	.22	277	270(t)	-.16		248.5(t)	-.51	232	230(p)	.05	227
CpApA	284	1.12	275.5	261	-2.99	244.5	236.5(p)	.38		225(p)	.65	221.5
CpApU	283.5	.85	274	261	-1.85	233	225(p)	.40	220			
CpApC	287.5	1.31	276	262.5	-2.42	233	226(p)	.77	219			
CpApG	289	.68	279.5	267.5	-1.60	231	225(p)	.25	220.5			
CpUpA	288.5	1.00	276.5	261.5	-1.55		225(p)	-.20				
CpUpU	287.5	1.36	276.5	260	-2.06		227.5(p)	-.09		217(t)	-.47	
CpUpC	289	1.59	277	260	-1.94		227.5(p)	-.46		218.5(t)	-.62	
CpUpG	290	1.09	279	266.5	-1.41		227.5(p)	-.18				
CpCpA	291	1.55	280	265.5	-2.04	228.5	226(p)	.23	223			
CpCpU	290.5	2.04	279.5	265	-2.21		226(p)	-.26				
CpCpC	291	2.35	281.5	267.5	-2.40		242.5(t)	-1.56		226(p)	-.40	
CpCpG	291.5	1.47	282.5	268.5	-1.78		242.5(t)	-.91		226(p)	-.22	
CpGpA	292.5	.43	277	257.5	-.97		237.5(p)	-.40		227.5(p)	-.34	
CpGpU	292.5	.62	281.5	270	-.57		255(p)	-.45		240(t)	-.62	
CpGpC	292.5	.92	279	257.5	-1.02		237.5(t)	-.71	222	215(p)	.30	
CpGpG	292.5	.44	282	251.5	-.93		227.5(p)	-.20				
GpApA	280	.91	272	260	-2.32	248	238.5(p)	.88		227.5(p)	.44	222
GpApU	277.5	.82	269	258.5	-1.21	241	230(p)	.24	221.5			
GpApC	285	.86	273.5	260	-1.67	237	227.5(p)	.55	219			
GpApG	287.5	.13		280(p)	.07	278	267(t)	-.66	246.5	240(p)	.15	225
GpUpA	290	.12	274	262.5	-.14	257.5	252.5	.09	248	240(t)	-.16	226
GpUpU	285	.52	275.5	262.5	-.63		250(P)	-.39	231	225(p)	.11	222.5

TABLE V (Cont.)

Compound	λ_p	$[\phi_p]$	λ_o	λ_t	$[\phi_t]$	λ_o	$\lambda_{(p\ or\ t)}$	$[\phi]$	λ_o	$\lambda_{(p\ or\ t)}$	$[\phi]$	λ_o
GpUpC	288	.70	277.5	265	-.52		254(p)	-.40		225(p)	-.26	
GpUpG	291.5	.23	283.5	272.5	-.47	263	255(p)	.38	244.5	237.5	-.23	227.5
GpCpA	290	.69	276.5	262	-1.17	232	225	.50	218			
GpCpU	290	1.17	277	260	-1.49	229.5	221(p)	.24				
GpCpC	292	1.49	280	267.5	-1.26		242.5(t)	-.90	225	222.5	.04	
GpCpG	292.5	.64	282	260	-.75	230	220(p)	.26				
GpGpA	275	.64	263.5	251	-.81	232	230(p)	.03	229			
GpGpU	282.5	.21		266.5(p)	.33	237.5	246(t)	-.56		227.5(p)	-.04	
GpGpC	292.5	.46		277.5(p)	.41	265	250(t)	-.81	226	220(p)	.08	
GpGpG	272.5	.58	259.5	250	-.89	234	230(p)	.17	226			

Note: t, p, and o refer to trough, peak, and cross-over, respectively.

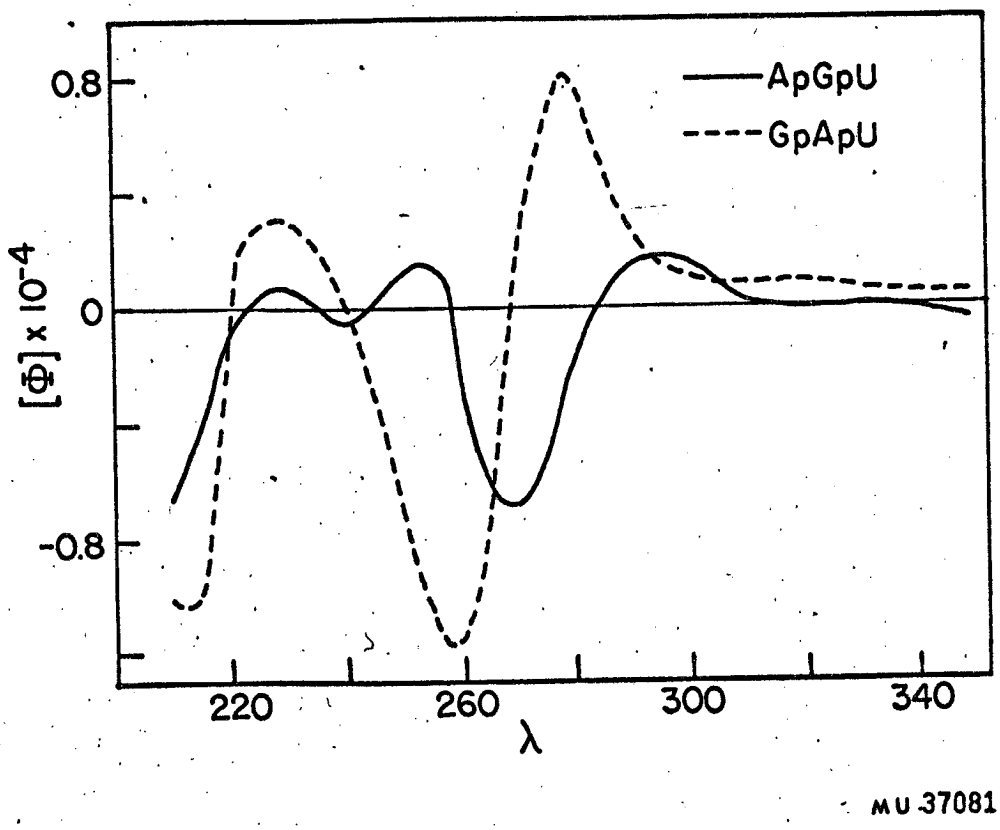


Figure 28. Sequence dependence of optical rotatory dispersion.

——— ApGpU.

- - - - - GpApU

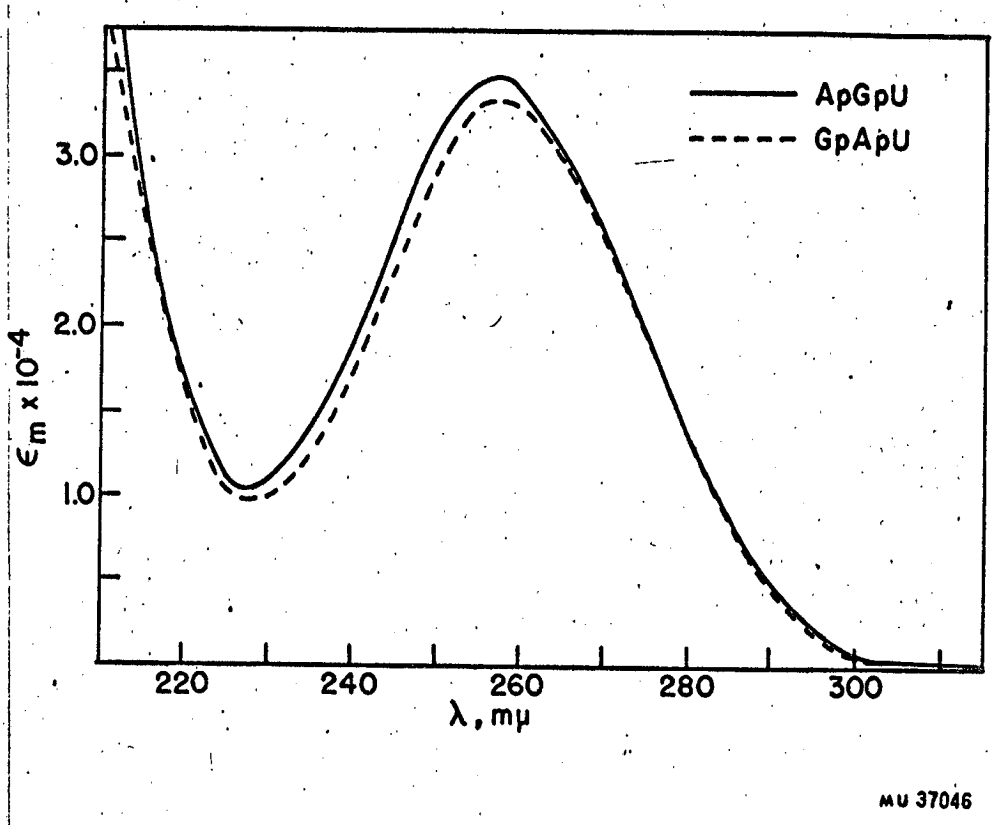


Figure 29. Sequence dependence of absorption spectra.

——— ApGpU

– – – – GpApU

We have also tabulated calculated results for the extinction coefficients and hypochromicities for the 64 trimers. These are shown in Table VI. Up till now the most widely used method of guessing at the extinction coefficient for an oligomer has been to add up the monomer absorption spectra. As our previous results have shown, nearest neighbor calculations usually do far better. Thus, this table of extinction coefficients should be of great assistance to the experimentalist who needs to know the amount of material he is using, but does not have enough trinucleoside diphosphate to determine its extinction coefficient by hydrolysis.

Several assumptions had to be made to obtain the results shown in Tables V and VI. All of our previous calculations involved an oligomer with either A or G as the middle base, but we do not see why this should effect the applicability of the calculations. The more serious problem is that only 15 of the 16 dinucleoside phosphates have been measured by Warshaw.[252] Since we had no experimental ORD for GpG we had to calculate it, and this was done by using our experimental data for the trinucleoside diphosphates GpGpC and GpGpU. The ORD of GpG is given by either of the following two equations.

$$[\phi]_{GpG} = 1/2(3[\phi]_{GpGpC} + [\phi]_{G} - 2[\phi]_{GpC})$$

$$[\phi]_{GpC} = 1/2(3[\phi]_{GpGpU} + [\phi]_{G} - 2[\phi]_{GpU})$$

There are significant differences in the ORD estimated for GpG by these two methods, and so we decided to average the two results. Whether this

TABLE VI

Calculated Residue Extinction and Hypochromicity at 260 mμ for 64 Trinucleoside Diphosphates

Compound	$\epsilon_{260} \times 10^{-4}$	Sum of $\epsilon_{260} \times 10^{-4}$ of monomers	% H(260)
ApApA	3.94	4.62	14.7
ApApU	3.60	4.07	11.5
ApApC	3.30	3.80	13.2
ApApG	3.70	4.23	12.5
ApUpA	3.87	4.07	4.9
ApUpU	3.37	3.52	4.3
ApUpC	3.13	3.25	3.7
ApUpG	3.41	3.68	7.3
ApCpA	3.48	3.80	8.4
ApCpU	3.00	3.25	7.7
ApCpC	2.80	2.98	6.0
ApCpG	3.16	3.41	7.3
ApGpA	3.87	4.23	8.5
ApGpU	3.47	3.68	5.7
ApGpC	3.09	3.41	9.4
ApGpG	3.51	3.84	8.6
UpApA	3.66	4.07	10.1
UpApU	3.32	3.52	5.7
UpApC	3.02	3.25	7.1
UpApG	3.42	3.68	7.1
UpUpA	3.43	3.52	2.6
UpUpU	2.93	2.97	1.3
UpUpC	2.69	2.70	0.4
UpUpG	2.97	3.13	5.1
UpCpA	3.10	3.25	4.6
UpCpU	2.62	2.70	3.0
UpCpC	2.42	2.43	0.4
UpCpG	2.78	2.86	2.8
UpGpA	3.37	3.68	8.4
UpGpU	2.97	3.13	5.1
UpGpC	2.59	2.86	9.4
UpGpG	3.01	3.29	8.5
CpApA	3.30	3.80	13.2
CpApU	2.96	3.25	8.9
CpApC	2.66	2.98	10.7
CpApG	3.06	3.41	10.3
CpUpA	3.09	3.25	4.9
CpUpU	2.59	2.70	4.1
CpUpC	2.35	2.43	3.3
CpUpG	2.63	2.86	8.0

TABLE VI (Cont.)

Compound	$\varepsilon_{260} \times 10^{-4}$	Sum of $\varepsilon_{260} \times 10^{-4}$ of monomers	% H(260)
CpCpA	2.80	2.98	6.0
CpCpU	2.32	2.43	4.5
CpCpC	2.12	2.16	1.9
CpCpG	2.48	2.59	4.2
CpGpA	3.15	3.41	7.6
CpGpU	2.75	2.86	3.8
CpGpC	2.37	2.59	8.5
CpGpG	2.79	3.02	7.6
GpApA	3.72	4.23	12.1
GpApU	3.38	3.68	8.2
GpApC	3.08	3.41	9.7
GpApG	3.48	3.84	9.4
GpUpA	3.59	3.68	2.4
GpUpU	3.09	3.13	1.3
GpUpC	2.85	2.86	0.3
GpUpG	3.13	3.29	4.9
GpCpA	3.12	3.41	8.5
GpCpU	2.64	2.86	7.7
GpCpC	2.44	2.59	5.8
GpCpG	2.80	3.02	7.3
GpGpA	3.53	3.84	8.1
GpGpU	3.13	3.29	4.9
GpGpC	2.75	3.02	8.9
GpGpG	3.17	3.45	8.1

discrepancy is due to experimental error is not known. But we are fairly certain that even if GpG were experimentally available, its ORD might be substantially different from the result calculated by the above method. This is due to the fact that the homo-oligomers containing several G's aggregate in solution even at low concentration.[175,132] Even GpG apparently exists as a high molecular weight aggregate at optical concentrations.[14] Thus a measurement of the

ORD of GpG in solution would yield the properties of the aggregate. But several G's in a larger oligomer will probably cause little aggregation. Thus we need to know the ORD of unaggregated GpG to predict the contribution of GpG to the ORD of a trimer such as GpGpA.

For the ORD on the other 15 dinucleoside phosphates we used the data of Warshaw.[252] The ORD data used for the monomers were the average of the nucleoside and nucleotide properties cited previously. All of these calculations were performed by the computer program discussed in Appendix 2. The data tabulated lists only major peaks and troughs. Small shoulders and wiggles are omitted.

We assume, for the moment, that the ORD calculated for the 64 trimers at pH 7 is as close to all of the experimental curves as it was to the seven we measured. In principle, all of the trinucleoside diphosphates should have different ORD curves, but in practice how many of the trimers could we distinguish from, say, their sequence isomers? Qualitatively, we can attempt to answer this question by looking at the ORD curves for various sequence isomers, but this is not very satisfactory. The 64 trimers can be divided into 20 different compositions. Four of these, the homotrimers contain only one oligomer. Four more comprise those trimers in which no two bases are the same. Each of these has six sequence isomers. The 12 remaining groups of type A_2B have three sequence isomers each. How many of these groups contain isomers which can be distinguished by an ORD measurement? To answer this question we have used a method worked out by Lee, McMullen, Brown, and Stokes for the analysis of mixtures of monomers.[124]

Whether we can identify an oligomer by measuring its ORD is analogous to whether we can quantitatively analyze for its presence in a mixture of all the isomers being considered. Obviously, if we can do the latter we can also do the former. The more direct question to ask is whether we can quantitatively analyze the presence of an oligomer in all the possible mixtures which contain it plus one other sequence isomer. But thus far our approach has been limited to the first type of analysis.[124] This sets a lower limit on our ability to distinguish a trimer. Experimental inaccuracy is considered by comparing our ability to analyze the oligomer in a mixture with our ability to determine its concentration when it is pure. We shall summarize the theory in the form that has recently been used.

Assume there is an n component mixture with concentrations c_j, where $j = 1$ to n. We make r measurements at wavelength λ_i ($i = 1$ to r) of an optical property, e_{ij} ($j = 1$ to n; $i = 1$ to r), that obeys Beer's law. The property of the mixture is just

$$E_i = \sum_j c_j e_{ij} , \quad \text{or in matrix notation,} \quad \underset{\sim}{E} = \underset{\approx}{e}\,\underset{\sim}{c} .$$

To solve these equations, we first multiply by the transpose, $\underset{\approx}{\tilde{e}}$, so that $\underset{\approx}{\tilde{e}}\underset{\approx}{e}$ is a square matrix. We then multiply both sides by the inverse of $\underset{\approx}{\tilde{e}}\underset{\approx}{e}$. The result is

$$\underset{\sim}{c} = (\underset{\approx}{\tilde{e}}\underset{\approx}{e})^{-1}\,\underset{\approx}{\tilde{e}}\,\underset{\sim}{E} , \quad \text{or} \quad c_j = \sum_i m_{ji} E_i$$

where m_{ji} is defined by this equation. Lee et al. show that if there is the same experimental error at every wavelength, σ_E, the error in measur-

ing the jth component is given by[124]

$$\sigma_{ij} = P_j \sigma_e \qquad \text{where } P_j = \left(\sum_i^n (m_{ji})^2 \right)^{1/2} .$$

If there is only a single component, we know that $c_j = E_i/e_{ij}$, and thus

$$P = \left(\sum_i^n (e_{ij})^2 \right)^{-1/2} .$$

This will give the minimum possible value of the error.

The quantity of interest is the ratio, R, of the error in analyzing an oligomer in a mixture to the error in analyzing the pure oligomer.

$$R = P_j/P$$

The magnitude of R is an indication of the distinguishability of the oligomer. If R is 1, the oligomer spectra or ORD are completely resolved from one another. If R is ∞, the oligomer properties are either identical or linearly dependent. Lee et al., use a value of R of about 10 as the limit of when a mixture can be analyzed.[124] Recent work suggests that this is too pessimistic.[138]

Using a computer program, Tinoco calculated the value of R for each of the mixtures of trimers discussed above.[238] The input for this program was the nearest neighbor ORD calculations. The results are shown in Table VII. Using the arbitrary and very conservative estimate that an R of 5 is the limit of resolution, we find that a total of 43 trimers can be resolved from a mixture of their sequence isomers. The remaining 21 have an average R greater than 5. Raising the level of permitted R to 10 would lead to a total of 49 distinguishable trimers out of the 64. But

TABLE VII

Identification of Trinucleoside Diphosphates by ORD

Easily Resolved			Poorly Resolved		
Mixture	Component	R*	Mixture	Component	R*
AG_2	GGA	1	UCG	GUC	5
	GAG	2		CGU	6
	AGG	2		UGC	7
	average	2		GCU	8
CG_2	GGC	2		UCG	9
	GCG	2		CUG	10
	CGG	2		average	8
	average	2	A_2U	AUA	4
UG_2	UGU	1		UAA	10
	GCU	2		AAU	12
	UGG	2		average	9
	average	2	AUC	AUC	6
U_2G	UGU	1		CUA	6
	UUG	3		CAU	12
	GUU	3		UAC	14
	average	2		UCA	14
A_2G	AGA	2		ACU	16
	GAA	2		average	11
	AAG	4	UC_2	CUC	11
	average	3		CCU	16
C_2G	CCG	3		UCC	17
	CGC	4		average	15
	GCC	5	U_2C	UUC	12
	average	4		UCU	13
ACG	AGC	3		CUU	21
	CGA	3		average	15
	CAG	4			
	ACG	5			
	GAC	5			
	GCA	6			
	average	4			
A_2C	ACA	3			
	CAA	5			
	AAC	6			
	average	5			
AU_2	UUA	4			
	UAU	5			
	AUU	5			
	average	5			
AUG	GUA	3			
	AUG	5			
	UGA	5			
	UAG	6			
	AGU	6			
	GAU	7			
	average	5			
AC_2	CCA	4			
	CAC	5			
	ACC	7			
	average	5			

*R is the ratio of the ease of analyzing for a component alone to analyzing for its presence in a mixture.

in view of the uncertainty in our calculated values, it is best in this preliminary study to be fairly conservative.

To show how the calculated values of R compare with our intuition, we have reproduced the ORD curves of several sets of sequence isomers in Figures 30 through 33. Figure 30 shows the set of isomers U_2G, which have an average R of 2 and thus can be easily distinguished. But the set of three isomers U_2A, shown in Figure 31, are difficult to distinguish. It can be seen that they have very similar shapes, and ApUpU and UpUpA appear to be close to linearly dependent. The average R for this system is 9, which confirms our suspicions. The ORD curves shown in Figures 32 and 33 represent the sets of 6 sequence isomers AUG and ACG respectively. In both of these cases, each of the curves seems to be reasonably different from any of the others. The values of R for these two systems are 5 and 4, and thus we can expect that both sets of isomers could easily be distinguished by measuring the ORD. Thus we can expect ORD in practice to be quite a powerful tool in determining the sequence of oligonucleotides.

12. Predicted Spectral Ratios of Trimers

The most common way of characterizing oligonucleotides has been the use of ratios of absorbance at various wavelengths. This may be a practice inherited from the biochemist, who is interested in separating mixtures of various types of components. Thus, the most commonly used ratio, A_{280}/A_{260}, is the same used by the biochemist to determine how much nucleic acid is contaminating his protein, or vice versa. But there is no particular reason why this spectral ratio or any of the

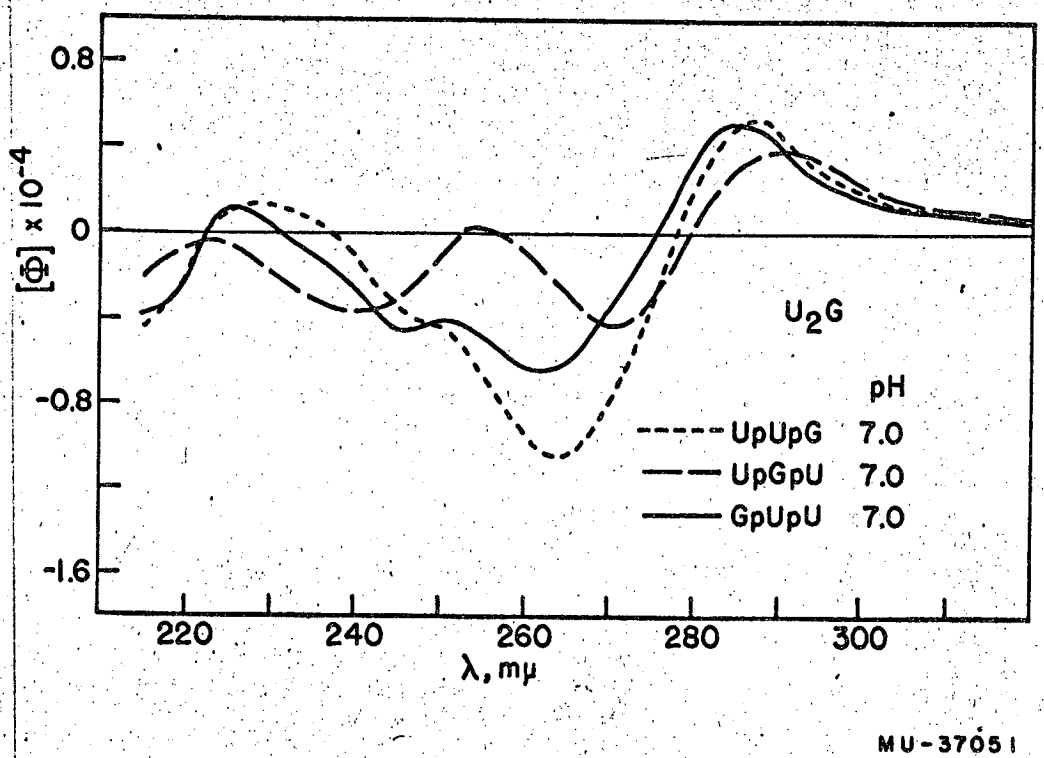


Figure 30. Sequence dependence of optical rotatory dispersion calculated by nearest neighbor methods. Three sequence isomers of composition U_2G.

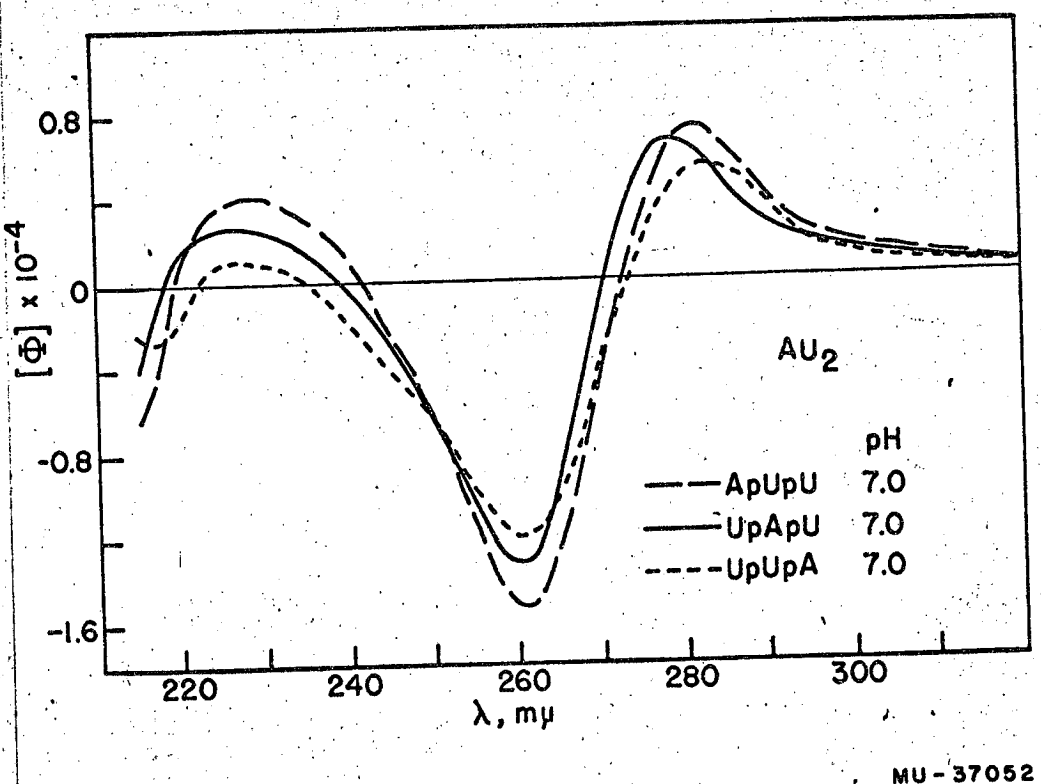


Figure 31. Sequence dependence of optical rotatory dispersion calculated by nearest neighbor methods. Three sequence isomers of composition AU_2.

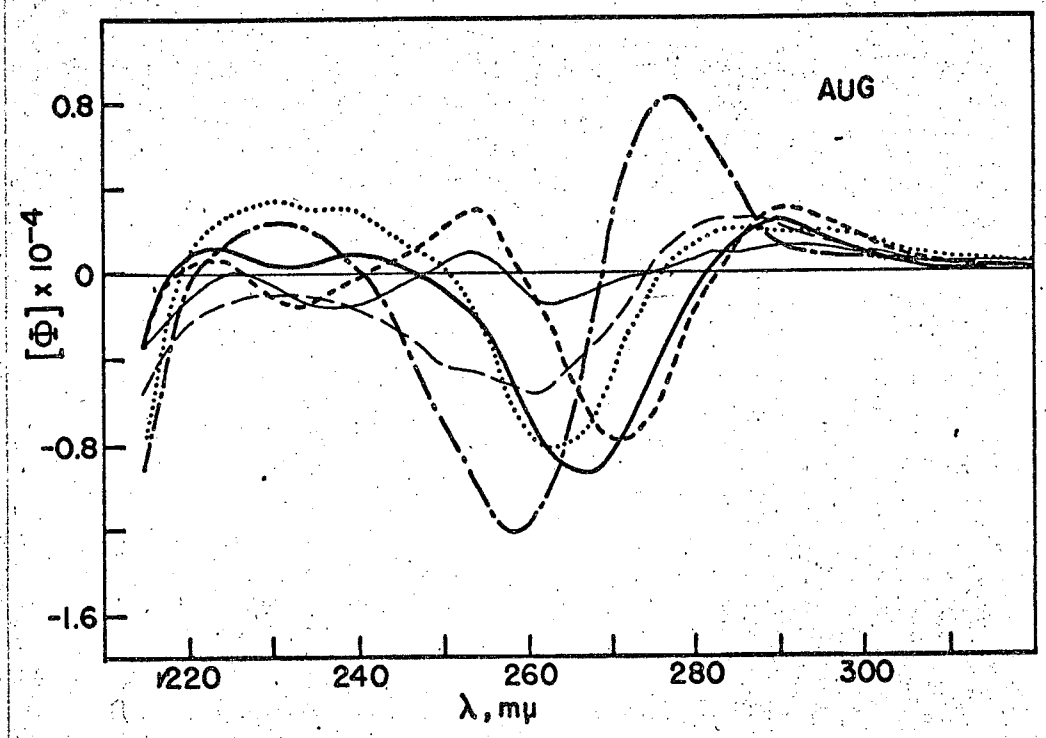


Figure 32. Sequence dependence of optical rotatory dispersion calculated by nearest neighbor methods. Six sequence isomers of composition AUG.

ApUpG
GpApU
UpApG
GpUpA
UpGpA
ApGpU

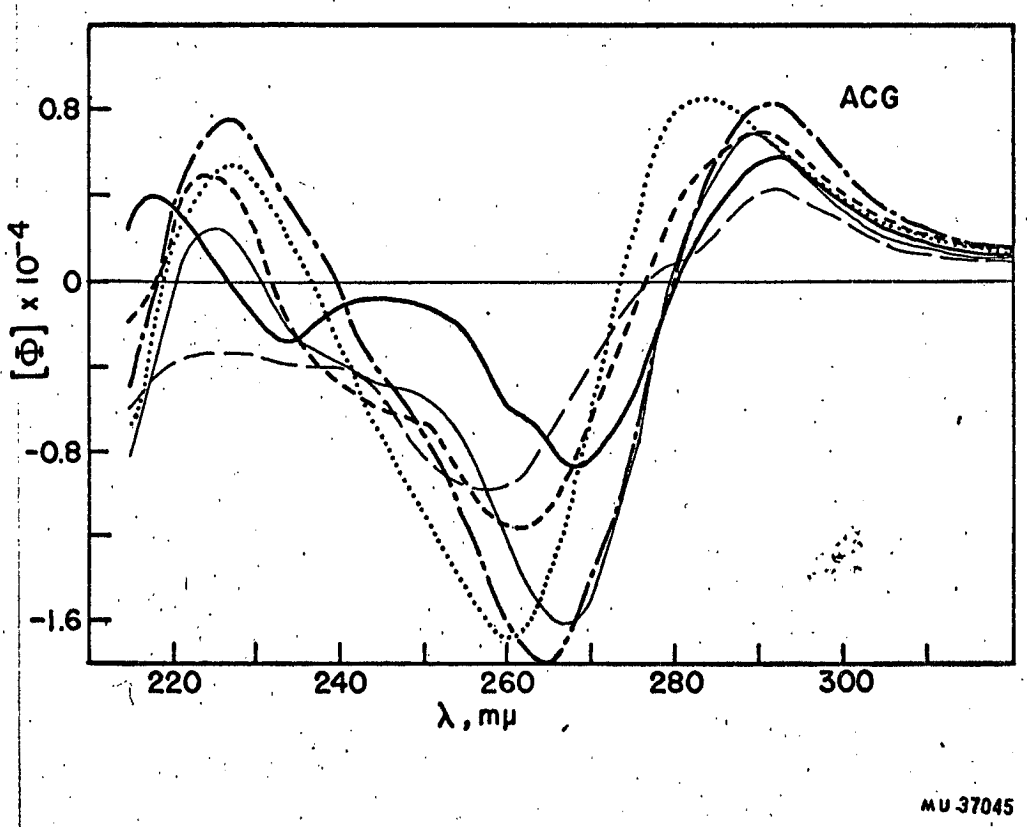


Figure 33. Sequence dependence of optical rotatory dispersion calculated by nearest neighbor methods. Six sequence isomers of composition ACG.

GpApC
ApCpG
ApGpC
CpApG
CpGpA
GpCpA

other popular ones, 250/260, 270/260, and 290/260 should be particularly useful in discriminating among the various possible oligonucleotides. In fact, it is our personal opinion that the wavelengths traditionally chosen are not the most useful ones. Nevertheless, since there is a large body of spectral ratios in the literature, we felt it would be worthwhile to calculate these ratios by the nearest neighbor methods outlined previously.

For a really accurate comparison of calculated and experimental values, we should have used the averages of several sets of dimer spectra, but only one determination of the extinction of these wavelengths was conveniently available to us.[252] The values for the monomers were taken from Pabst circular OR-10,[166] and fortunately, these spectral ratios do not change much from nucleoside to nucleotide.

Experimental values for the 250/260, 280/260, and 290/260 ratios of various trinucleoside diphosphates or trinucleotides are compared with the calculated values at three pH's in Table VIII. While there are some values for other wavelengths in the literature, we decided to confine our attention to these three, since they are the most frequently reported. A brief perusal of the values shown in Table VIII should leave one with several qualitative conclusions. The accuracy of determining these ratios leaves much to be desired; large deviations are observed from one worker to the next. The agreement between experimental and calculated curves is usually fairly good, and in a majority of cases is better than the agreement between experimental values and the sum of the monomers.

TABLE VIII

Spectral Ratios of Trinucleoside Diphosphates

Compound	pH	Ref	A_{250}/A_{260}			A_{280}/A_{260}			A_{290}/A_{260}		
			Observed	Calc'd†	Monomer sum	Observed	Calc'd†	Monomer sum	Observed	Calc'd†	Monomer sum
ApGpU	1	*	.838	.839	.848	.431	.398	.419	.215	.171	.191
"	1	222	.86	.839	.848	.42	.398	.419	.208	.171	.191
"	1	121	.89	.839	.848	.42	.398	.419		.171	.191
"	7	*	.923	.903	.897	.404	.368	.389	.138	.119	.105
"	7	222	.92	.903	.897	.38	.368	.389	.122	.119	.105
"	7	121	.84	.903	.897	.41	.368	.389		.119	.105
"	11.5	*	.845	.849	.830	.393	.380	.353	.095	.100	.048
"	11	222	.85	.849	.830	.36	.380	.353	.061	.100	.048
"	10	121	.84	.849	.830	.41	.380	.353		.100	.048
GpGpC	1	*	.827		.850	.885		.979	.627		.732
"	1	222	.83		.850	.85		.979	.592		.732
"	7	*	1.035		1.043	.659		.748	.299		.297
"	7	222	1.05		1.043	.64		.748	.286		.297
"	11.5	*	.871		.883	.694		.709	.197		.169
"	11	222	.88		.883	.65		.709	.156		.169
GpGpU	1	*	.869		.892	.594		.598	.372		.370
"	1	222	.88		.892	.61		.598	.348		.370
"	7	*	1.029		1.042	.584		.591	.252		.213
"	7	222	1.02		1.042	.62		.591	.220		.213
"	11.5	*	.864		.873	.574		.545	.118		.094
"	11	222	.89		.873	.65		.545	.105		.094
ApGpC	1	222	.81	.801	.805	.69	.710	.744	.440	.449	.500
"	1	121	.79	.801	.805	.74	.710	.744		.449	.500
"	7	222	.94	.931	.931	.47	.460	.514	.178	.178	.171
"	7	121	.95	.941	.931	.51	.460	.514		.178	.171
"	11	222	.87	.879	.838	.48	.494	.514	.136	.137	.114
"	10	121	.95	.879	.838	.49	.494	.514		.137	.114

TABLE VIII (Cont.)

Compound	pH	Ref	A_{250}/A_{260} Observed	A_{250}/A_{260} Calc'd†	A_{250}/A_{260} Monomer sum	A_{280}/A_{260} Observed	A_{280}/A_{260} Calc'd†	A_{280}/A_{260} Monomer sum	A_{290}/A_{260} Observed	A_{290}/A_{260} Calc'd†	A_{290}/A_{260} Monomer sum
ApCpG	1	121	.81	.835	.805	.73	.684	.744		.420	.500
"	7	121	.92	.930	.931	.55	.463	.514		.182	.171
"	10	121	.92	.860	.838	.56	.466	.514		.141	.114
ApApC	1	*	.789	.801	.768	.575	.556	.548	.334	.302	.306
"	1	222	.78	.801	.768	.54	.556	.548	.348	.302	.306
"	1	121	.80	.801	.768	.54	.556	.548		.302	.306
"	7	*	.827	.810	.801	.423	.371	.326	.131	.091	.071
"	7	222	.83	.810	.801	.39	.371	.326	.099	.091	.071
"	7	121	.81	.810	.801	.36	.371	.326		.091	.071
"	11.5	*	.860	.833	.801	.431	.383	.326	.154	.093	.071
"	11	222	.83	.833	.801	.39	.383	.326	.099	.093	.071
"	10	121	.83	.833	.801	.37	.383	.326		.093	.071
ApApU	1	*	.836	.839	.811	.308	.278	.267	.080	.046	.040
"	1	222	.83	.839	.811	.28	.278	.267	.054	.046	.040
"	1	121	.65	.839	.811	.26	.278	.267		.046	.040
"	7	*	.813	.803	.779	.326	.273	.225	.062	.025	.017
"	7	222	.81	.803	.779	.28	.273	.225	.034	.025	.017
"	7	121	.82	.803	.779	.24	.273	.225		.025	.017
"	11.5	*	.832	.830	.794	.295	.246	.198	.048	.035	.011
"	11	222	.83	.830	.794	.25	.246	.198	.026	.035	.011
"	10	121	.84	.830	.794	.26	.246	.198		.035	.011
GpApU	1	*	.844	.855	.848	.417	.416	.419	.195	.194	.191
"	1	222	.86	.855	.848	.40	.416	.419	.191	.194	.191
"	1	121	.85	.855	.848	.36	.416	.419		.194	.191
"	7	*	.881	.886	.897	.417	.405	.389	.133	.128	.105
"	7	222	.90	.886	.897	.40	.405	.389	.122	.128	.105
"	7	121	.86	.886	.897	.44	.405	.389		.128	.105

TABLE VIII (Cont.)

Compound	pH	Ref	A_{250}/A_{260} Observed†	A_{250}/A_{260} Calc'd†	A_{250}/A_{260} Monomer sum	A_{280}/A_{260} Observed	A_{280}/A_{260} Calc'd†	A_{280}/A_{260} Monomer sum	A_{290}/A_{260} Observed	A_{290}/A_{260} Calc'd†	A_{290}/A_{260} Monomer sum
GpApU	11.5	*	.786	.842	.830	.391	.380	.353	.077	.081	.048
"	11	222	.85	.842	.830	.37	.380	.353	.054	.081	.048
"	10	121	.85	.842	.830	.43	.380	.353		.081	.048
ApUpG	1	121	.87	.856	.848	.49	.399	.41)		.191	.191
"	7	121	.91	.881	.897	.48	.374	.389		.086	.105
"	10	121	.91	.840	.830	.48	.359	.353		.076	.048
GpApC	1	222	.81	.814	.805	.69	.738	.744	.449	.492	.500
"	7	222	.93	.903	.931	.50	.524	.514	.193	.203	.171
"	11	222	.86	.845	.838	.51	.533	.514	.139	.145	.114
ApGpA	1	121	.86	.863	.871	.39	.352	.352		.172	.171
"	7	121	.91	.893	.896	.36	.308	.306		.108	.083
"	10	121	.91	.827	.821	.38	.385	.292		.065	.037
ApApG	1	121	.87	.884	.871	.34	.369	.352		.180	.171
"	7	121	.91	.940	.896	.33	.357	.306		.106	.083
"	10	121	.89	.854	.821	.33	.355	.292		.063	.037
UpApU	1	121	.82	.820	.779	.27	.326	.325		.060	.043
"	7	121	.83	.790	.761	.27	.299	.299		.039	.029
"	10	121	.83	.809	.798	.27	.245	.244		.041	.018
ApUpU	1	121	.82	.815	.779	.34	.311	.325		.052	.043
"	7	121	.80	.788	.781	.33	.288	.299		.036	.029
"	10	121	.80	.834	.798	.34	.249	.244		.047	.018
ApUpC	1	121	.76	.753	.725	.64	.613	.658		.340	.351
"	7	121	.85	.813	.786	.44	.399	.423		.099	.093
"	10	121	.83	.865	.806	.43	.397	.404		.121	.092

TABLE VIII (Cont.)

Compound	pH	Ref	A_{250}/A_{260} Observed	A_{250}/A_{260} Calc'd†	A_{250}/A_{260} Monomer sum	A_{280}/A_{260} Observed	A_{280}/A_{260} Calc'd†	A_{280}/A_{260} Monomer sum	A_{290}/A_{260} Observed	A_{290}/A_{260} Calc'd†	A_{290}/A_{260} Monomer sum
ApCpU	1	121	.73	.739	.725	.64	.614	.658		.318	.351
"	7	121	.78	.806	.786	.42	.403	.423		.103	.093
"	10	121	.80	.816	.806	.41	.396	.404		.106	.092
GpCpU	1	121	.77	.769	.764	.83	.799	.899		.483	.583
"	7	121	.91	.975	.938	.61	.572	.660		.202	.217
"	10	121	.91	.917	.854	.60	.589	.635		.149	.151

†Using equation 7

*This work

It is possible to put the above qualitative conclusions on a more quantitative basis. The average difference of experimental values of ratios from the average experimental value is tabulated below for three spectral ratios. This should be compared with the average differences between the experimental ratios and those calculated by nearest neighbor methods or by summing the monomer spectra. It can be seen that one does significantly better by using nearest neighbor calculations than by simply adding the monomers. In light of the experimental inaccuracy, however, it is doubtful whether this is worth the effort, in most cases.

	250/260	280/260	290/260
Average experimental error	.012	.018	.012
Average difference between calculated and experiment	.019	.032	.019
Average difference between monomer sum and experiment	.028	.038	.029

Looking only at the experimental curves, one can see that the range of 250/260 spectral ratios is very small, and thus is of almost no use in typing oligomers. The 280/260 ratios differ more, but the 290/260 ratio especially at pH 1 seems to be the most useful in distinguishing various oligomers.

One conclusion which can be drawn from Table VIII is that there will often be very poor agreement between experimental spectral ratios and those calculated from a sum of the monomer spectra. In this light it is interesting to note that Lane has stated that the 250/260,

280/260, and 290/260 spectral ratios of his 2'-O-methyl trinucleotides correspond very closely to the values calculated from the sum of the normal mononucleotides.[117] Two of the oligomers he studied are AxpGxpCp and GxpAxpUp (where x means 2'-O-methyl). From Table VIII it can be seen that several of the spectral ratios, particularly 290/260, for the corresponding normal oligomers, differ from the ratios calculated from a sum of the monomers by much more than the average experimental error discussed above. Thus, for example, the 280/260 ratio of GpApU at alkaline pH averages .419 among three determinations, and the sum of the monomers predicts only .353 for an average difference of .066. The experimental 290/260 ratio of ApGpC at pH 1 is .440, while the sum of the monomers predicts .500. Thus, if we can assume that Lane used the same frame of reference in deciding that his spectral ratios agreed well with the monomers, we can say that his trimers resemble the monomers much more than the normal trimers do. Is this reasonable? We think it is, since the extra methyl groups can restrict the number of possible conformations of the ribose phosphate backbone which permit stacking. Thus, the ΔS of stacking should be more negative, and these methylated trimers should be less stacked than in the normal analogues. This agrees with the above spectral evidence, though we will feel much more comfortable when ORD data for the methylated oligomers becomes available.

13. Other Applications of Nearest Neighbor Calculations

In this section we shall discuss several additional ways in which nearest neighbor calculations can help us to analyze oligonucleotides.

Previous examples have all dealt with measurements made on systems with time independent behavior. But ORD measurements can also be of use in following the kinetics of reactions. For example, consider the hydrolysis of a tetranucleotide by venom phosphodiesterase. The two sequence isomers ApUpApA and ApApUpA are expected to have the same ORD and other optical properties, since they have the same types of nearest neighbors; but if we allow exoenzyme degradation to occur, the intermediates in the reaction are different. For example, after one bond is broken we would have

$$\text{ApApUpA} \longrightarrow \text{ApApU} + \text{A}$$

$$\text{ApUpApA} \longrightarrow \text{ApUpA} + \text{A}$$

Once the reaction has gone to completion, the ORD of the two reaction mixtures should be the same, since the products in both cases are A + 2pA + pU. Thus, if we measure the ORD as a function of time, we can expect that at intermediate times in the reaction it will be different depending on which tetramer we have.

The concentrations of intermediate chain length oligomers in an exoenzyme digest can easily be calculated using the methods worked out in Chapter II. By combining these results with the ORD of each of the intermediates, calculated by the methods described in this chapter, we can compute the ORD as a function of time. The time dependence of the ORD at 260 mμ is compared for two choices of kinetic parameters, in Figures 34 and 35. The steady state case of the random model was used to calculate the ORD of ApApUpA and ApUpApA reaction mixtures, shown in Figure 34. It can be seen that the time dependence of the ORD is

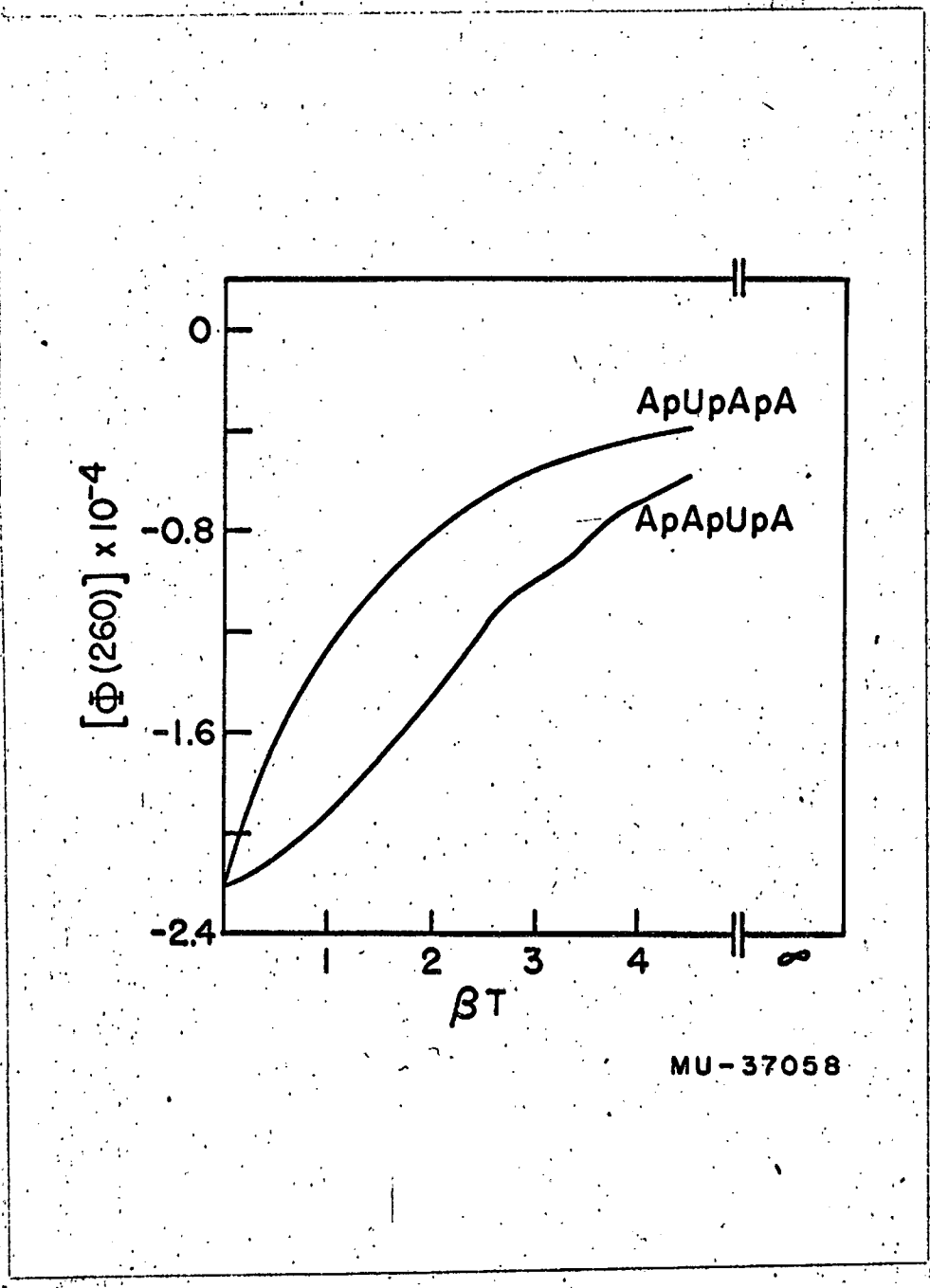


Figure 34. Time dependence of the optical rotation at 260 mμ of a tetranucleotide in the presence of venom phosphodiesterase. Calculated by nearest neighbor methods using the steady state case of the random enzyme model discussed in Chapter II.

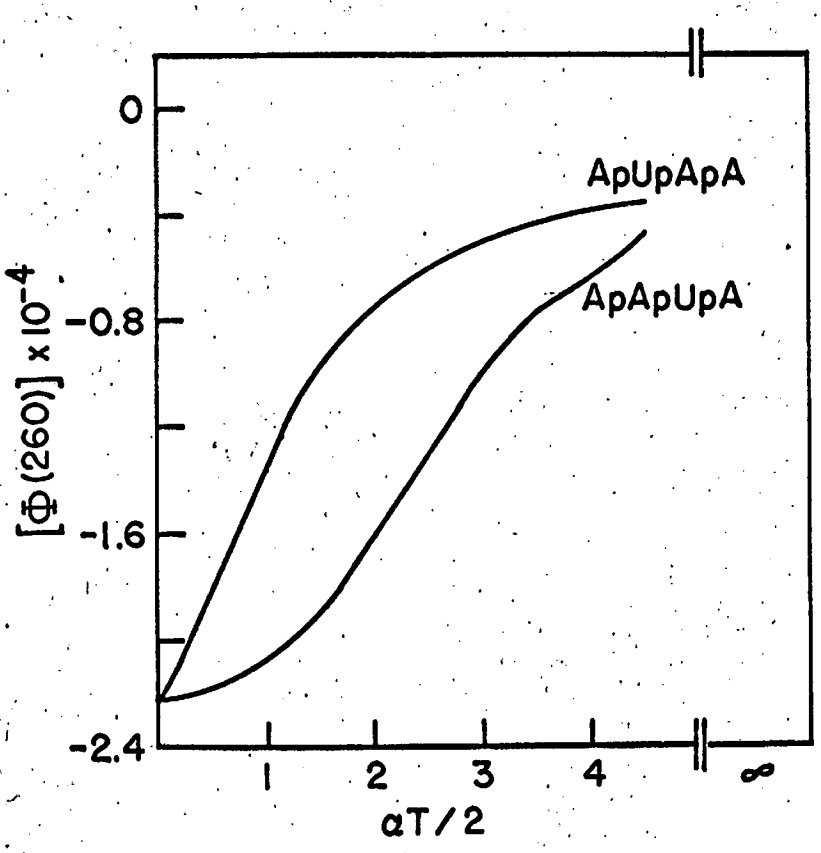


Figure 35. Time dependence of the optical rotation at 260 mμ of a tetranucleotide in the presence of venom phosphodiesterase. Calculated by nearest neighbor methods using the optimal case of the random enzyme model discussed in Chapter II.

very different for the two oligomers. Thus we can determine sequence by measuring the kinetics of exoenzyme degradation without having to perform an analysis of the monomer released. This is a great advantage, since the problems of sampling are avoided. Calculations for the same two oligomers using the optimal case of the random model are shown in Figure 35. As expected, the differences between the two sequence isomers are even larger here. Thus it should be very easy to distinguish between sequence isomers in this manner. And while the oligomer would be degraded, the amount of material needed is much less than in other methods.

One complication could arise which might make the above method of sequencing difficult. There is much evidence that when nucleotides are mixed with nucleases substantial spectral changes occur.[51] If this happens with exoenzymes, and if the spectral changes are accompanied by large changes in ORD, the analysis will become much more complicated. It is impossible to say whether the result will lead to more or less precision in sequence determination. But it will certainly lead to a powerful tool for studying the properties of these enzymes.

If polynucleotide phosphorylase is used as an exonuclease, the above analysis must be modified. When highly pure, this enzyme cannot degrade dinucleoside phosphates. Thus the ORD of the final reaction mixture will be different for two sequence isomers such as ApApUpA and ApUpApA, since in the former case ApA will remain while in the latter ApU is undegraded. This will provide additional information, and should make terminal sequence determination by these enzymes more useful. Kinetic changes in ORD will certainly help to determine the sequence of many oligomers which have identical nearest neighbor pairs.

Endoenzyme degradation can not, in general, be used to help determine the sequence of oligomers in which the same nearest neighbors can give rise to several sequences. For example, complete hydrolysis of GpGpCpG with T1 RNAase will yield the dimer CpG as well as the monomer G. Treatment of the sequence isomer GpCpGpG will yield the same final products if terminal phosphates are ignored. Likewise, treatment of both tetramers with pancreatic RNAase will yield GpC and G. Thus, there is no additional optical information to be gained by performing an endoenzyme hydrolysis.

The change in the ORD of a polynucleotide upon endoenzyme degradation can provide some sequence information.[238] For example, if an RNA is digested to completion with T1 RNAase, the ORD change for this process should be proportional to the number of GpA, GpU, GpC, and GpG bonds broken. Thus, measuring this change will tell us what fraction of the resulting oligomers end in 5' A, U, or C. In practice, great sensitivity may be needed to be able to detect this change accurately, since for an average RNA it can represent at most 1/4 of the ORD of an average polymer.

Thus far, all of the ORD measurements we have discussed have been made at room temperature. Current experiments by Davis will soon provide us with a knowledge of the temperature dependence of the ORD and absorption of all 16 dinucleoside phosphates and the four monomers.[37] It will be of great interest to see how well nearest neighbor calculations can predict the temperature dependence of the optical properties of oligonucleotides. We certainly expect that at high temperature the ORD of all oligomers should resemble the monomers, and this is the case for the systems already studied. Although calculated and experi-

mental results may agree at high and low temperature, there is no guarantee that the shape of the temperature transition of dimers is the same as trimers. It is possible that next-nearest neighbor interactions may be more important for thermodynamic properties than they are for ORD or absorption. But if we are confident that the ORD can be predicted for a given conformation from the dimers, analysis of discrepancies between experimental results and calculations at intermediate temperatures can provide clearer thermodynamic information.

Other techniques such as solvent perturbation and spectrophotometric titrations have yet to be studied in detail using the ORD of oligonucleotides. This should provide a fertile field for future work. It will be of great interest to learn just how general is the nearest neighbor treatment of oligonucleotides.

In summary, we have developed a method by which the optical properties of oligonucleotides can be easily predicted. We have shown that the conformation of bases in trinucleoside diphosphates in solution is the same as the conformation of these bases in the component dinucleoside phosphates. This is an important first step in understanding the conformation of polynucleotides in solution.

IV. OPTICAL PROPERTIES OF POLYNUCLEOTIDES

1. General Considerations of RNA Structure

In the past chapter we intentionally confined our attention to the optical properties of single stranded oligonucleotides. Now we shall treat the complications which arise when an RNA strand has the capability of forming conformations which include double or multiple stranded sections. As in the past, our general approach will be to try and understand the properties of polynucleotides in terms of the properties of oligonucleotides.

It has been known since the late 1950's that a single strand is a poor model for the conformation of a natural RNA in solution. Even RNA molecules which were proven to be one continuous strand showed properties which indicated a complex secondary and tertiary structure.[217] In this section we shall review much of the evidence that has led to our current conceptions of the conformation of RNA in solution. Discussion of some recent papers which are especially pertinent to the conformation of TMV RNA and sRNA is postponed to later sections.

X-ray diffraction techniques provided the clue to the understanding of the conformation of DNA.[255] But attempts to apply this method to structure determination of RNA have not progressed at as fast a rate. The problem is that most RNA structures are not as regular as DNA, and thus the diffraction patterns are more blurred, and their analysis more difficult. Novertheless, good fiber patterns have been obtained from some RNA's, and approximate structures have been determined. The most

well understood RNA is obtained from reovirus, and has a molecular weight of about 10^7. Langridge and Gomatos[118] have shown that at least 50% of this RNA is a double strand, and their data are consistent with 90 to 100% double strands in the fiber, but the patterns are too blurred to be sure. The structure resembles the A form of DNA (low humidity) with a 30 Å helix repeat and 10 residues per turn. The planes of the base pairs are tilted 75 to 80 degrees to the helix axis. But unlike DNA, when the humidity is raised there is no structure transition to a conformation in which the base planes are perpendicular to the helix axis. The difference between the X-ray fiber structure of double strand RNA and DNA is probably not due to the replacement of T by U, because Langridge and Marmur have shown that PSB2 DNA, which contains U instead of T, has the same three dimensional structure as normal DNA.[119] Thus the difference between DNA and RNA can only be attributed to the presence of the 2' OH group in the latter. This is supported by the fact that very large differences are observed between the properties of poly dA and poly A.[28]

Studies by Tomita and Rich have lead to the conclusion that fibers of TMV RNA, disoriented Wound Tumor Virus, ribosomal RNA and a random copolymer of A and U (1:1) all have small double stranded segments.[239] In addition, the undenatured Wound Tumor Virus RNA shows a well resolved X-ray fiber pattern similar to reovirus. Fibers of the acid form of poly A have been shown to consist of double strands in which both strands are parallel, unlike DNA.[181] Intensive efforts have resulted in the preparation of only one crystalline RNA, and this has a dubious pedigree.[216] This RNA, which was originally thought to be sRNA, has now

been shown to be degraded rRNA. While X-ray results on this low molecular weight RNA are not clear, they indicate that it consists of small double strand sections with a structure similar to reovirus. Thus all the X-ray results agree that under some conditions RNA can be partially double stranded.

Reovirus RNA contains two complementary double strands, and thus is unlike most of the other RNAs which have been studied. It is interesting to note how similar the properties of this double strand are to the more familiar DNA. The melting curve of reovirus RNA is very sharp.[78] Krug, Gomatos and Tamm have recently shown that this RNA will act as a template for DNA primed RNA polymerase.[113] By analyzing the nearest neighbor frequencies of the product formed they have been able to show that the two strands of reovirus RNA are antiparallel. This received a strong confirmation when the same workers showed that the RNA will also act as a template for DNA polymerase, resulting in the synthesis of double stranded antiparallel DNA.[77] This leads us to suspect that in solution this RNA must have a structure extremely close to that of DNA. In contrast to reovirus RNA, most RNAs are very poor templates for polymerases.[76] Thus they are probably not nearly as perfect double strands as reovirus.

Of course, one must always be careful in extrapolating the results of fiber or crystal structures to predict the structure of a polymer in solution. Fortunately, there have been several studies of the small angle X-ray scattering of RNA solutions, and these are fairly consistent with the more detailed information available from fibers. Witz, Hirth and Luzzati have looked at rRNA, TMV RNA, and TYMV RNA at salt

concentrations which range from 0.15 to 10^{-3} molar.[259] All three are quite similar and show a mass per unit length similar to DNA in solution. But these molecules are not as rigid, and are more compact. The structures suggest that the base planes are close to perpendicular to the helix axis. These results are interesting since, as Witz, Hirth and Luzzati point out, TYMV RNA can have a maximum of 78% Watson-Cuick base pairs. But it appears to be just as ordered as TMV RNA, which is potentially capable of forming more A-U and G-C hydrogen bonds than TYMV RNA. Thus the authors suggested that other types of base pairs must play an important role in stabilizing RNA double strands.

The other double strand RNA which has been studied by small angle X-ray methods is the acid form of poly A.[137] It was found that the mass per unit length is slightly less than DNA. This agrees with the fiber studies discussed earlier.[181] Thus there is much evidence that double stranded RNA occurs in solution. It is hoped that these studies will soon be extended to other ordered polynucleotides.

Some of the earliest evidence that RNA could form double strands came from the study of the UV spectra of mixtures of synthetic polynucleotides. Solutions of mixtures of poly A and poly U show sharp melting transitions at pH 7.[224] Mixing experiments have led to the idea that a 1 to 1 double strand complex is stable in the absence of magnesium ion. If Mg^{++} is added, the triple strand poly (A+2U) becomes the stable form under the range of concentrations studied.[224] However, the interpretation of UV mixing curves is not always clear. Drobnik and Kleinwachter have shown that the results can strongly depend on the choice of wavelength.[46] Also, mixing curves cannot distinguish among helices of the type AnBn, where n has any integral value.[30]

Fresco has claimed that poly (A+U) is not the equilibrium form of a 1 to 1 mixture of poly A and poly U even at as low salt concentrations as 0.1 molar.[66] Fortunately, the discrepancy between his interpretation and the earlier results of Steiner and Beers has been partially clarified by the recent work of Miles and Frazier.[153] They have used IR spectroscopy, which has the advantage that the absorption peaks corresponding to all of the expected components in a mixture of poly A and poly U are well resolved. Thus the composition of each mixture can be clearly determined. Miles and Frazier show that at neutral pH in 0.1 M Na^+, poly (A+U) is the stable complex in mixtures containing less than 0.5 mole fraction U. Under these conditions, poly (A+2U) is stable above 0.67 mole fraction U. At intermediate compositions, both forms are in equilibrium. The addition of Mg^{++} has no effect except around 0.6 mole fraction U.[153] By an extension of this work, Miles now suspects that in the triple strand complex the two strands of poly U are parallel to each other and antiparallel to poly A.[152]

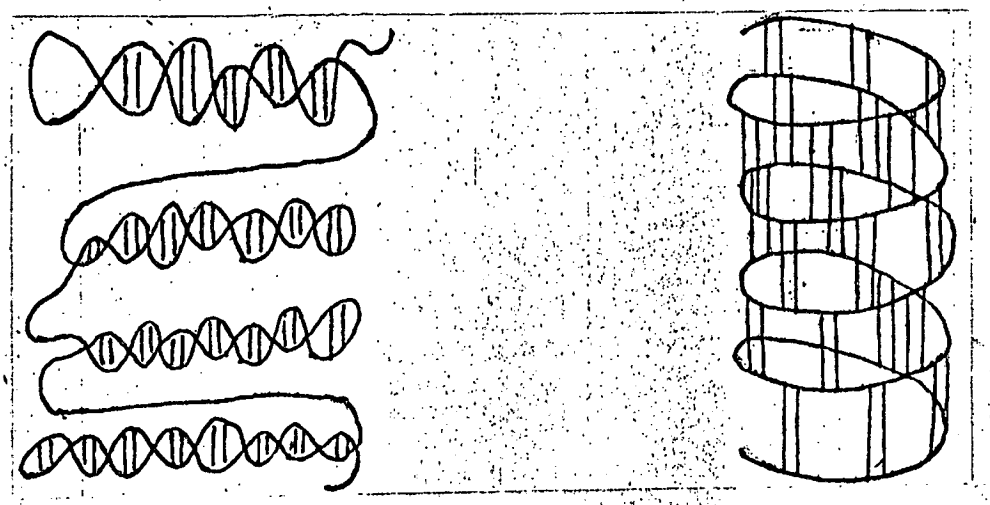
It has been more difficult to obtain satisfactory results with mixtures of poly C and poly G. In the past, satisfactory preparations of poly G have been hard to come by. But recent work by Pochon and Michelson has shown that poly C and poly G form a stable 1:1 complex at room temperature in 0.1 molar salt.[174] This complex is very stable and in 0.002 M Na^+ shows no hyperchromic change up to 100°C. If short chains of G are used instead, one observed a 2G + C triple strand.[174] This seems to explain the earlier results of Fresco.[66] Thus, G-C and A-U pairs can exist in polyribonucleotides in solution. But these experiments do not prove that these are the only possibilities.

A third regular polyribonucleotide which has been studied in solution is poly r(AU), a copolymer of alternating sequence.[29] This polymer shows a sharp melting transition which occurs at a higher temperature than the corresponding poly d(AT). The melting transition is reversible, even when measured by viscometry, which indicates that poly r(AU) is probably a single strand. The secondary structure arises from the strand folding back on itself to form a double strand.

The above experiments indicate that A-U and G-C pairs are important in RNA structure at neutral pH. When the pH is lowered, there is a tendency for these pairs to be broken apart and others to be formed between the protonated bases. This is readily observed by studying the titration of the homopolymers. Poly A in 0.1 M KCl picks up 1/2 an equivalent of protons absuptly at pH 5.9 to form a double strand helix with A-A base pairs. Unlike DNA, this helix is destabilized by salt. Thus in 0.01 M salt, the transition occurs at pH 6.3.[224] Hartman and Rich have found similar results for poly C. At pH 5.7, this polynucleotide takes up 1/2 equivalent of H^+ and forms a double strand with C-C base pairs.[85] These transitions of poly A and poly C are accompanied by large changes in the ORD.[54,95] Thus, it should be kept in mind that to explain the optical properties of RNA at pH's lower than neutrality one may have to consider conformations involving protonated bases.

Thus far, most of the experiments which we have been discussing are concerned with what is conventionally called secondary structure. The recent review by Spirin cites much evidence that RNA has tertiary structure as well.[217] Electron micrographs of TMV RNA or E. coli rRNA

deposited from solutions of low ionic strength show that these molecules are long rods. Under these conditions TMV RNA is found to have a diameter of 30 Å and a length of up to 2500 Å. This is a thicker rod than would be expected for a simple double strand helix. These results are confusing, and it is possible that they represent experimental artifacts. Other data reported by Spirin suggests that simple double strands are an oversimplified picture of RNA conformation. The dichroism of the 260 absorption band of RNA oriented in an electric field is positive. RNA is expected to orient with the long axis parallel to the field. Since the transitions responsible for the 260 transition are $\pi \rightarrow \pi^*$, polarized in the plane of the bases, this means that the base planes are parallel to the long axis of the molecule. To explain this unexpected result, Spirin has proposed the structure shown below, a rod composed of small parallel helices.[217] Another possibility exists if one allows the presence of hydrogen bonded base pairs between parallel RNA strands. Then it is possible to form a helix with a large diameter containing base pairs parallel to the helix axis. This structure, which would also show positive dichroism, is shown below for comparison.



small helices

large II helix

At the present time, both of these structures can be regarded as highly speculative. In most of this chapter we will restrict ourselves to considerations of RNA secondary structure, for which there is much more conclusive evidence.

2. Evidence for A-U and G-C Pairs

Much of the past evidence for the details of RNA conformation comes from the study of the temperature dependence of UV spectra. Interpretation of RNA melting curves is complicated by the fact that both single and double strand conformations change with increasing temperature. Since RNA melting curves tend to be sharper than transitions expected for pure single strands, there has been a tendency to ascribe all of the effects observed to the breaking of hydrogen bonds.[66] This error should be kept in mind while evaluating the results presented below.

RNA melting profiles are not so sharp as DNA. This led to the idea that RNA is composed of short double helical segments spaced by more random conformations. The breadth of the melting profile is thought to arise from the continual melting of segments of different length and stability. Recently, Fresco has shown that some purified sRNA's have a two phase melting curve.[66] Each is much sharper than the relatively broad melting one sees in unfractionated yeast sRNA. This strongly suggests that sRNA contains double (or multiple) stranded sections, since these are expected to have a much sharper melting transition than single strands.

There has been much detailed analysis of melting transitions in an attempt to resolve some of the details of RNA conformation. Fresco and his collaborators have analyzed their data in three different ways to provide evidence on the types of base pairing in RNA. The first approach is to compare T_m (the temperature at which half of the possible change in absorbance has occured) with the base composition of a set of RNA's. Fresco claims that X_gX_c, X_aX_u, etc., where X is mole fraction, correspond to the probability of each type of base pair occuring, and thus he plotted all of these possible products against T_m.[194] Good correlation occurs only between T_m and X_aX_u or X_gX_c. But there are many sRNA's among the 10 RNA's he considered, and the above plot is only true for infinite polymer. To correct this, Fresco also sought correlations between T_m and X_gX_c / $(X_gX_c + X_aX_u)$. These correlations occur, but all that this can tell one is that interactions between G and C are stronger than interactions between A and U. It would be interesting to see a plot of the minimum value of X_a or X_u against T_m, since this corresponds to the maximum possible number of interactions of a given type, but Fresco does not plot this directly. Thus, the above evidence suffers from the weakness that the correct plots were not made. What one really would like to know is the average number of interactions expected for a given base composition in a finite polymer. This could then be compared with T_m. Unfortunately, to compute this average, some assumptions have to be made about the nature of the interactions. But it is clear that the average will involve both terms of the type X_G or X_C minimum, and X_gX_c. At present, it can be concluded from these data only that A-U and G-C pairs may have some importance.

The second method used by Fresco and his collaborators to explore RNA secondary structure is the analysis of difference spectra.[68] They attempt to decompose the spectral change upon heating into contributions from A-U and G-C pairs, using the following equation,

$$A(\lambda)/c = X_{AU}\Delta\epsilon_{AU}(\lambda) + X_{GC}\Delta\epsilon_{GC}(\lambda)$$

where A is the absorbance, c is the concentration, and X is the mole fraction of base pair. $\Delta\epsilon_{AU}$ and $\Delta\epsilon_{GC}$ are hypothetical changes in extinction for forming an A-U or G-C pair. Fresco obtained $\Delta\epsilon_{AU}$ from the spectral change upon forming poly(A+U) from poly A and poly U. But for $\Delta\epsilon_{GC}$ Fresco and his collaborators use the change in extinction upon heating poly GC (48:52) random copolymer. This is a very bad choice, since it ignores the fact that this copolymer may have G-G and C-C as well as G-C pairs. Thus, while it is possible to fit the change in absorbance of an RNA using Fresco's $\Delta\epsilon_{AU}$ and $\Delta\epsilon_{GC}$, one cannot interpret a good fit as meaning that there are only A-U and G-C pairs. Felsenfeld and Cantoni have tried to correct for the above error in $\Delta\epsilon_{GC}$ by using the change in extinction upon forming oligo G poly C pairs.[56] But there is still one error left in this method, since it assumes that the interactions leading to hypochromicity come from interstrand interaction only. In fact, there are also major contributions from intrastrand stacking. By including nearest neighbor interactions in their analysis of DNA spectra, Felsenfeld and Hirschman have shown that quadratic terms will appear in the above equations.[57] At some wavelengths these can become large. We hope that this more complete treatment can soon be extended to RNA spectra. Until then one cannot be completely satisfied with the conclusion that difference spectra show only A-U and G-C pairs.

The third method used by Fresco to analyze melting curves is to plot T_m against wavelength.[66] This turns out to show a minimum T_m at the wavelengths where the A-U difference spectrum is largest. This shows that A-U pairs are weaker than other interactions, but it still is not conclusive evidence for the existence of only A-U and G-C pairs in RNA.

One of the most direct methods for studying hydrogen bonded interactions is hydrogen exchange. While this method was long plagued by uncertainties, recent technical improvements have led to new respectability and the ability to study exchange reactions with much shorter half lives. In their recent study of hydrogen exchange in DNA, Printz and von Hippple assume that all non hydrogen bonded exchangeable protons have half lives too short to permit observation.[171] This permits the decay curves to be extrapolated to zero time and yield an estimate of the number of hydrogen bonds. These extrapolated results for calf thymus and T4 DNA are within 5% of the number of hydrogen bonds calculated for completely complementary DNA with known base composition. Thus one can feel confident that the exchange one is observing is really due to hydrogen bonded atoms. In native DNA, the kinetics of exchange are simple 1st order, which suggest that all the hydrogen bonds are approximately equivalent. In denatured DNA, the decay curves are more complex, and extrapolation to zero time implies that only 54% of the hydrogen bonds remain. Since the hypochromicity of this denatured DNA is 75% of the native value, estimates of helicity based on hypochromism are now open to question. The discrepancy between these two values presumably comes from the contribution of single strand conformations to the hypochromicity.

The concurrent studies of Englander and Englander on RNA provide an interesting comparison with the DNA results.[53] These authors propose that exchange can occur only when the hydrogen bonded pair breaks and swings out of the helix, and this is the rate limiting step. Extrapolation to zero time for the sRNA's they studied show about 1.1 hydrogen bond per nucleotide pair. The rate of exchange is very sensitive to Mg^{++} concentration. If Mg^{++} is present, the exchange rates are independent of salt concentration. But in the absence of Mg^{++}, strong dependence on salt concentration is observed. Englander and Englander conclude that sRNA is highly helical, but does not have a static structure.

We are now ready to state the postulates about RNA structure which most workers take for granted. As the above discussion indicates, we are by no means convinced that these postulates are valid. But they form an effective working hypothesis. Most attempts to understand the conformation of RNA have been based on these six assumptions of Fresco, Alberts and Doty.[67]

1. Only A-U and G-C pairs are important.
2. At 25°C and moderate ionic strength, the minimum stable double strand helix is about 4 residues long.
3. Double strands are formed by hairpin turns which contain a non hydrogen bonded loop of at least 3 residues.
4. Looping out of additional bases can occur to enable more hydrogen bonds to be formed. Without looping out, one can expect that only 20% of the residues of a random chain are capable of forming A-U and G-C pairs, within the above restrictions.

5. Most stable regions involve interactions between adjacent sections of the polynucleotide chain. Interactions between distant residues restrict the entropy of the intermediate chain.

6. All double strands are antiparallel, like those in DNA.

It should be realized that most of these postulates are untested. We have cast doubt on postulate 1 earlier in this chapter. Postulates 2, 3, and 4 are, at best, approximations. Since stacking energy will be lost when a hairpin turn is made the length of a stable hairpin may well depend on the sequence of non hydrogen bonded residues. The only consideration that should be made in deciding the conformation of RNA is to minimize the free energy of the solution. Thus, under some circumstances it is quite possible for distant sections of the polymer chain to interact. The last postulate is true in long double strand reovirus RNA, but we see no reason why it is necessary for short double strand segments. We hope that the methods we shall describe later in this chapter will eventually be used to test the validity of these postulates. Until such time, we will not let them restrict our conception of the possible conformations of an RNA strand in solution.

3. Interactions Between Oligonucleotides and Polymers

To understand the properties of double strand RNA it will first be helpful to study small oligomers which are capable of forming double strand complexes. Since the difficulty of preparing an oligonucleotide greatly increases with chain length, it is desirable to find the smallest compounds which can still form double strand complexes like longer poly-

mers. In this section we shall review some of the experimental evidence for interaction between oligonucleotides or between oligomers and polymers.

There is much evidence that at sufficiently high concentration mononucleosides or mononucleotides will interact in solution. But much of this evidence, discussed in the last chapter, indicates that these interactions are vertical stacking and thus correspond to single strand polymers. One exception is pG, which forms a gel which contains hydrogen bonded complexes of G as well as stacked rings.[70] But there seems to be no well established evidence that purines will specifically interact with pyrimidines in aqueous solution to form hydrogen bonded pairs. On the contrary, the recent NMR evidence of Schweizer, Chan and Ts'o suggests that purines interact with pyrimidines in solution by vertical ring stacking. The interaction between two purines seems to be slightly stronger than between a pyrimidine and a purine.[200] Thus to find evidence for double strand interactions in solution, the dimer is the shortest chain length worthwhile considering.

In crystals, the interaction between bases or nucleosides to form complementary pairs is well established. Haschemeyer and Sobell have found that a crystal grown from a 1:1 mixture of dG and 5-bromo dC in aqueous solvent contains pairs of bases in which each G forms three hydrogen bonds with a C just as in DNA.[88] The same hydrogen bonds have been found for mixed crystals of derivatives of the bases C and G. In all cases, hydrogen bonded base pairs are reported. Mixed crystals of adenosine and 5-bromouridine also contain hydrogen bonded pairs, as do other derivatives of these bases or nucleosides. But in this case, the

A-U pairs are not always hydrogen bonded as in DNA.[87] In some cases the U pairs with the A through the amino group and the imidazole nitrogen. Thus, in this case, the A-U pair resembles the interaction between poly A and the second strand of poly U in poly(A+2U). Whether the occurence of this unusual hydrogen bonding is due to lattice forces is not known. But it is encouraging that purine pyrimidine interactions seem to be quite frequent in the solid state. Only one dinucleoside phosphate has been studied by the methods of X-ray crystallography.[203] This is Adenylyl(2'→5')uridine, which in principle could form an antiparallel double strand complex with two A-U pairs. But this was not observed in the crystal.

Interaction between oligomers on a chromatographic column might represent a situation halfway between the solid state and solution. Tuppy and Kuchler have bound bases covalently to a resin and used this to separate mixtures of other bases. They find that the base complementary to the one bound on the column is always retarded. Thus, through the magnification of chromatography, it has been possible to demonstrate that single bases can interact with each other to favor complementary pairs.[244] The polynucleotide cellulose columns devised by Gilham represent an extension of this idea. The success of these columns indicates that fairly specific interactions can occur between oligonucleotides and short homopolymers bound to a resin.[73]

There is much evidence that oligomers will interact with homopolymers in solution to form double or triple strand complexes. Most of this evidence comes from the study of UV spectra, though recently the methods of IR spectroscopy have also been used. The nature and stability of the

complexes formed depends, in general, on the concentration of the reactants, the composition of their mixture, the salt concentration, the absence or presence of Mg^{++}, and the pH. In Table I we have attempted to summarize the currently available data for oligomer-polymer complexes. It can be seen that in most cases complexes occur at concentrations not much higher than those ordinarily used for UV spectroscopy. The complex becomes more stable as the chain length of the oligomer is increased.

Lipsett, Heppel and Bradley have studied the interaction of adenylic acid oligomers with poly U under a variety of conditions.[133,134] In general, they find that with chain lengths of A greater than three, the presence of Mg^{++} strongly stabilizes the triple strand A+2U complex. The double strand formed in the absence of Mg^{+} has a T_m twenty degrees lower. The presence of a non-complementary base in oligomers such as $p(Ap)_3U$ lowers the melting temperature considerably. The presence of a terminal phosphate seems to have almost no effect. Thus it appears that the behavior of these oligomer-polymer complexes is a good model for the interaction between polymers. As Lipsett, Heppel and Bradley point out, one must view melting temperatures below 10°C with suspicion since it is known that here poly U begins to form a double strand complex itself.[134,129]

The effect of non-complementary bases on the interaction between oligomers and poly U has been studied recently in more detail by Bautz and Bautz.[6] They find that a terminal C in an oligomer of the type $(Ap)_nX$ does not destabilize the double or triple strand complexes formed nearly as much as a terminal G or U. However, random poly (AC) forms the weakest complex with poly U. This should serve to caution us that extrapolations from oligomer interactions to polymer interactions must be done cautiously.

TABLE I

Oligomer-Polymer Interactions

Oligomer	Polymer	Complex	Concentration† (mM of residue)	pH	Salt Concentration (molar)	Mg^{++} Conc (molar)	T_m °C	Reference
$(pA)_2$	U	1:1	.06	7.4	1 M NaCl	.001	7	134
$(Ap)_2$	"	1:1	.06	7.4	" "	.001	5	"
ApA	"	1:1	.06	7.4	" "	.001	6.2	"
$(pA)_3$	"	1:2	.06	7.4	" "	.001	17	"
$(Ap)_3$	"	1:2	.06	7.4	" "	.001	23	"
$(Ap)_2A$	"	1:2	.06	7.4	" "	.001	17.4	"
$(pA)_3$	"	1:2	.06	6.9	0.15 M NaCl	.05	17.6	"
$(pA)_3$	"	1:1	.065	7.2	0.10 "	none	2 - 4	"
$(Ap)_2A$	"	1:1	.08	7.5	0.50 "	none	16	6
$(Ap)_2A$	"	1:2	.08	7.5	none	.10	31	"
$(pA)_4$	"	1:2	.06	7.4	1 M NaCl	.001	29.5	134
$(Ap)_4$	"	1:2	.06	7.4	" "	.001	33.2	"
$(Ap)_3A$	"	1:2	.06	7.4	" "	.001	31.3	"
$(pA)_4$	"	1:2	.06	6.9	0.15 M NaCl	.05	32.2	"
$(pA)_3pU$	"	1:2	.06	6.9	" "	.05	11.4	"
$(pA)_4$	"	1:1	.065	7.2	0.10 "	none	11.5	"
$(Ap)_3A$	"	1:1	.08	7.5	0.50 "	none	27	6
$(Ap)_3C$	"	1:1	.08	7.5	" "	none	12	"
$(Ap)_3G$	"	1:1	.08	7.5	" "	none	8	"
$(Ap)_3A$	"	1:2	.08	7.5	none	.10	42	"
$(Ap)_3C$	"	1:2	.08	7.5	"	.10	26	"
$(Ap)_3G$	"	1:1	.08	7.5	"	.10	19	"
$(Ap)_3U$	"	1:1	.08	7.5	"	.10	19	"
$(pA)_5$	"	1:2	.06	7.4	1 M NaCl	.001	37	134
$(Ap)_5$	"	1:2	.06	7.4	" "	.001	40.5	"
$(pA)_4pU$	"	1:2	.06	6.9	0.15 M NaCl	.05	20	"
$(pA)_4pUp$	"	1:1	.06	7.4	1 M NaCl	.001	11.5	"
$(pA)_4pU$	"	1:1	.06	7.4	" "	.001	16.3	"
$(Ap)_4C$	"	1:2	.08	7.5	0.5 M NaCl	none	22	6

TABLE I (Cont.)

Oligomer	Polymer	Complex	Concentration† (mM of residue)	pH	Salt Concentration (molar)	Mg^{++} Conc (molar)	T_m °C	Reference
$(Ap)_4G$	U	1:1	.08	7.5	0.5 M NaCl	none	18	6
$(Ap)_4U$	"	1:1	.08	7.5	" "	"	16	"
$(Ap)_4C$	"	1:2	.08	7.5	none	.10	36	"
$(Ap)_4G$	"	1:2	.08	7.5	"	.10	34	"
$(Ap)_4U$	"	1:2	.08	7.5	"	.10	32	"
$(pA)_7$	"	1:2	.06	7.4	1 M NaCl	.001	49.7	134
$(pA)_9$	"	1:2	.06	7.4	" "	.001	54.6	"
pG	C	1:1	230.	7.8	0.2 M phosphate	none	13*	100
pG	"	1:2	163.	7.0	0.2 M cacodylate**	"	22.5*	"
GpG	"	1:1	.65	6.2	1 M NaCl	"	25.4	131
GpG	"	1:1	.063	7.4	0.2 M NaCl	"	14	"
GpG	"	1:1	.075	7.4	" "	"	22.5	"
GpG	"	1:1	.063	7.4	0.4 M NaCl	"	8.0	"
GpG	"	1:1	.063	7.4	1.0 M NaCl	"	6.4	"
GpG	"	1:2	36.	6.7	0.6 M cacodylate**	"	38.*	100
$(Gp)_2$	"	1:1	.80	6.2	0.2 M NaCl	"	16	131
$(Gp)_2$	"	1:2	4.2	6.2	" "	"	27.1	"
$(Gp)_2G$	"	1:1	.80	6.2	" "	"	45.6	"
$(Gp)_2G$	"	1:2	.063	6.2	" "	"	42.2	"
$(Gp)_2G$	"	2:1	.80	6.2	" "	"	25.5	"
$(Gp)_3$	"	1:1	.80	6.2	" "	"	45.0	"
$(Gp)_3$	"	1:2	.80	6.2	" "	"	45.0	"
$(Gp)_3$	"	2:1	.80	6.2	" "	"	24.7	"
$(Gp)_3$	"	1:1	160.††	8.3	0.13 M cacodylate**	"	34 - 37*	100
$(Gp)_2U$	"	1:1	.74	6.2	0.2 M NaCl	"	15	131
$(Gp)_4$	"	1:1	.80	6.2	" "	"	58.3	"

†Concentrations where not given in reference were estimated from high temperature absorbance of solutions.
*Estimated from shape of melting profile.
**Sodium salt.
††Estimate from path length of cell used.

The interactions between oligomers containing G and poly C are slightly more complex than those we have described above. Depending on the conditions used, Lipsett has found that it is possible to form either a 1:1 complex or either of the possible 2:1 complexes.[131] From the data shown in Table I it is readily apparent that complexes between G oligomers and poly C have much higher T_m's than complexes between the same chain lengths of A and poly U. This is consistent with the evidence presented earlier that the G-C pair is stronger than the A-U pair. Howard, Frazier, Lipsett and Miles have studied some of these interactions and, in addition, found that the mononucleotide d(pG) could form a double strand complex with poly C.[100] No similar interaction has been published for poly U and A. It would be interesting to see whether complexes between oligo G and poly C, or complexes between oligo U and poly A are very different from the complementary complexes discussed here. Unfortunately, we know of no published studies of this type.

4. Oligomer-Oligomer Interactions

In the past section we discussed some of the extensive evidence for the interaction between oligonucleotides and homopolymers in aqueous solution. But there is very little published data on the interaction of oligomers with oligomers. Most of the experiments which have been reported treat the aggregation of oligomers containing large amounts of G. It is not clear whether the strong tendency of G containing oligomers to aggregate is pertinent to the structure of RNA in solution. As we mentioned previously, most workers assume that only A-U and G-C pairs are important contributors to RNA conformation. But we feel that the effect of G-G pairs cannot be assumed to be negligible.

The only systematic study of the aggregation of G containing oligonucleotides has been carried out by Lipsett.[130] She finds that the aggregates are formed extremely slowly in solution, thus suggesting that they have very high molecular weight. The stability of the aggregates is very sensitive to the concentration of oligomer, and not so sensitive to the chain length, the pH, or salt concentration. These results are summarized in Table II. GpGpG readily aggregates at normal optical concentrations, and GpG shows a melting temperature of about 20°C at an O.D. of only 6. Interactions of oligomers containing several G's are often observed in chromatographic systems.[180] Thus, interactions between sequences containing 2 or more G's may play an important role in RNA structure. But these runs of G should occur fairly infrequently in a random RNA. Thus Watson-Crick type base pairing is probably more important.

The only information on the possibility of A-U hydrogen bonded pairs in oligomers in solution comes from the recent work of Thach and Sundararajan.[234] They report that $U(pU)_6(pA)_2$ has no hypochromicity in the range of 7° to 50°, but the oligonucleotide $U(pU)_6(pA)_4$ shows a temperature transition with a T_m of about 8°. The temptation is to say that this is due to the formation of a hairpin loop with 3 or 4 A-U hydrogen bonded pairs. But since studies of T_m against concentration have not been reported, we do not know whether this is due to inter or intramolecular effects.

From the above dearth of experimental evidence, it is obvious that a study of the interaction of oligonucleotides in solution is very worthwhile. This is one of the most direct approaches to ascertaining

TABLE II

Oligomer - Oligomer Interactions

Oligomer	Concentration† (mM of residue)	pH	Salt Concentration (molar)	T_m °C	Reference
$d(pG)_3$	.12*	6.8	0.25 M phosphate	58	175
$d(pG)_3$	.13*	6.55	0.25 M succinate	28	"
$d(pG)_4$	.11*	6.55	" "	47	"
$d(pG)_4$	.11*	6.8	0.2 M NaCl	36	"
GpG	.042	6.2	0.2 M cacodylate**	none	130
GpG	.46	6.2	" "	17.4	"
GpG	.432	6.2	1.0 M "	23.0	"
GpG	.042	7.2	0.2 M "	none	"
GpG	.378	7.2	" "	18.5	"
GpG	.042	8.8	" "	none	"
GpG	.378	8.8	" "	16.4	"
$(Gp)_2$	4.32	5.0	0.2 M acetate	26.2	"
$(Gp)_2$	4.32	6.2	0.2 M cacodylate	22.3	"
$(Gp)_2$	0.70	6.2	" "	none	"
$(Gp)_2$	any	7.4	" "	"	"
$(Gp)_2G$	.42	6.2	" "	23.6	"
$(Gp)_2G$	.09	7.4	" "	21.8	"
$(Gp)_2G$	.42	7.4	" "	24.3	"
$(Gp)_2G$	.042	8.8	" "	23.5	"
$(Gp)_2G$	.42	8.8	" "	28.4	"
$(Gp)_3$	.42	6.2	" "	23.7	"
$(Gp)_2U_p$	.49	6.2	" "	none	"
$U(pU)_6(pA)_2$	?	?	?	none	234
$U(pU)_6(pA)_4$	?	?	?	8	"

†Concentrations where not given in reference were estimated from high temperature absorbance of solutions.
*Estimated from shape of melting profile.
**Sodium salt.
††Estimate from path length of cell used.

the types of interactions that are important for RNA structure. Thus, it is of interest to determine what chain lengths are necessary before two complementary oligomers can form a stable double strand complex. It is to the experimenter's advantage to work with the shortest chain lengths feasible, since these are easier to prepare. In addition, intramolecular effects are less likely to confuse the analysis of interactions between, say, trimers or tetramers.

Since there was no experimental evidence for the interaction of complementary dinucleotides or trinucleotides, we decided to try and estimate the degree of association of these compounds at concentrations which can be conveniently reached experimentally. Consider the reaction between two dimers to form a double strand with two hydrogen bonded pairs. The ΔH for this process arises from two types of interactions. ΔH_s comes from the increased stacking interaction between the bases on the same strand. We define ΔH_s to contain the sum of single strand interactions of both strands. Thus, for the dimerization of oligomers containing n bases, there is a change in enthalpy of $(n-1)\Delta H_s$. The second interaction, ΔH_b, is due to the interactions of bases on one strand with bases on the other. This may include both hydrogen bond enthalpy and additional stacking between bases on one strand with bases on the other. As before, ΔH_b contains contributions from both strands. Thus, for an oligomer of chain length n we shall assume that there is a change in enthalpy of $n\Delta H_b$ upon interaction. This is certainly an approximation, but we are only interested in obtaining a crude estimate of the interaction between short oligomers. Thus we shall also assume that $\Delta H_s = \Delta H_b = \Delta H_n$. This is probably not a bad first assumption, and has been

used previously by Magee, Gibbs and Zimm.[240] The last assumption we shall make is that the entropy change, ΔS, is a constant, ΔS_r, times the number of residues of base pairs formed in the double strand.

With the above approximations we can write, for the reaction A + B = C, that $\Delta F = 3\Delta H_r - T2\Delta S_r$. A and B are dimers, and C is the double strand complex. If θ is the fraction of B that is in a double strand, and K is the equilibrium constant, we have

$$\theta = \frac{(C)}{(C) + (B)} \qquad K = \frac{(C)}{(A)(B)} \quad . \tag{1}$$

In this section we express concentration in moles of oligomer, not moles of residue. If we restrict ourselves to cases in which (A) = (B), we find that

$$K = \frac{\theta}{(B)(1-\theta)} \quad . \tag{2}$$

Equation (2) will permit us to calculate the melting temperature (temperature when $\theta = 1/2$) as a function of concentration, and ΔF.

To estimate the melting temperature of the interaction between two dimers, we need to know ΔH_r and ΔS_r. We can obtain those quantities from the experimental data on the interaction of homodimers with homopolymers. Magee, Gibbs and Nowell have solved the statistical mechanics of this problem.[139] We shall make use of their earlier approximate treatment which assumes that oligomers are tightly bound to the polymer.[140] There are no dangling ends. For the interactions of dimers with polymers, this is an excellent assumption.[231] Using the experimental T_m shown in Table I for ApA interacting with poly U, one can estimate

from the calculations of Magee, Gibbs and Zimm that $\Delta H/RT = -3.91$ per mole base pair.[140] This gives $\Delta H_r \approx -2.2$ kcal per mole. Thus, for an infinite polymer, one would estimate a total ΔH of about -4.4 kcal per base pair. This is a reasonable value, but smaller than the experimental results of Ross and Scruggs for the interaction of poly A and poly U.[184] A second possible choice of ΔH comes from the work of Applequist and Damle.[3] Using their "staggering zipper" model to fit the unpublished results of Fresco, Blake and Doty, they estimate a ΔH of about -8 kcal for the formation of double strand oligo A's. This would lead to a ΔH_r of about -4 kcal per mole. This, coincidently, seems to be a good upper estimate for the ΔH_r of forming double stranded Watson-Crick base pairs. Thus we shall estimate the T_m for the interaction of dinucleotides using both values of ΔH_r.

To obtain a value for ΔS_r, we must estimate the equilibrium constant for the interaction of ApA and poly U from the experimentally available melting temperature. As a first approximation, the binding of dimers onto a polymer can be looked upon as a simple Ising problem.[93] We shall assume that the types of interactions between oligomers and oligomers are the same as the oligomer polymer case. We find that each dimer gains 2 ΔH_b and 1 ΔH_s when it binds to the polymer. The loss in entropy is 2 ΔS_r per dimer. But, in addition, the dimers bound to the polymer can interact with each other through vertical stacking. The strength of this interaction should be ΔH_r. It is shown by Hill that for N particles bound onto an Ising lattice containing M sites

$$\frac{(N_{11})(N_{00})}{(N_{01})^2} = \frac{e^{-w/kt}}{4} \qquad (3)$$

where (N_{11}) is the number of times particles occupy adjacent sites, (N_{01}) is the number of times particles are next to an unoccupied site, and (N_{00}) is the number of unoccupied sites which are not adjacent to an occupied site.[93] These three quantities are related by two constraints, $2N = 2N + N_{01}$, and $2(M-N) = 2N_{00} + N_{01}$. If, for example, the lattice is half occupied ($\theta = 1/2$) these constraints, combined with equation (3) lead to the result that

$$(N_{11}) = \frac{Ne^{-w/2kt}}{1 + e^{-w/2kt}} \qquad (4)$$

As w/2kt becomes -2 or less, N_{11} approaches N. Thus all the bound particles occupy sites adjacent to other bound particles. Therefore, we can look upon the process as an all or none binding of a set of particles onto the lattice. Since w/2kt is approximately -2 or less for the cases we are considering, it will not be a bad approximation to treat the binding of ApA onto poly U as an all or none phenomenon.

For the process $nA + B = C$, we can write that $\Delta F = n\Delta H_b + (n-1)\Delta H_s - T2n\Delta S_r$. Using the same assumptions as before, we can simplify this result to $\Delta F = 2n(\Delta H_r - T\Delta S_r)$. With K and θ defined as before, we can write for the above interaction between dimers and polymer

$$\theta = \frac{(C)}{(C) + (B)} \qquad K = \frac{(C)}{(B)(A)^n} \,. \qquad (5)$$

These equations can be combined to give the result that

$$K = \frac{\theta}{(A)^n\ (1-\theta)} \qquad (6)$$

Now we are ready to calculate T_m for two dimers interacting. From the results shown in Table I, we see that ApA and poly U show a T_m of about 280°K. We shall assume that the poly A contained approximately 100 residues. Magee, Gibbs and Newell have shown the results are rather insensitive to the chain length of the polymer.[139] The total concentration of poly U plus ApA was $.06 \times 10^{-3}$ molar in nucleosides. Thus the initial concentration of ApA is 1.5×10^{-5} molar. When half of the dimer is bound (at T_m) the concentration of free ApA is thus $.75 \times 10^{-5}$ molar. Using equation (6), we calculate K and find that $\Delta F/RT_m = -543$. Using the experimental result for T_m leads to the following equation.

$$\Delta S = \frac{\Delta H + (1086)(280)}{(280)} \text{ eu} . \qquad (7)$$

We have shown that $\Delta H \cong 200\ \Delta H_r$. Thus, by using our two limiting values of ΔH_r in equation (7), we have two possible values for ΔS. We find

ΔH_r	ΔS_r
-2.2 kcal	4.78 eu
-4.0 kcal	17.74 eu

These values probably represent reasonable extremes for the thermodynamics of the interactions.

Using equation (2) we can compute the values of K at the melting temperature for the interaction of two dimers. These are then combined with the result that $\Delta F = 3\Delta H_r - T_m\, 2\Delta S_r$. The calculated T_m's for several concentrations, and our two choices of ΔH_r, are shown on the following page.

Concentration		ΔH_r -2.2 kcal	ΔH_r -4.0 kcal
(A)=(B)	10^{-4} molar	224°K	217°K
	10^{-3} molar	265°K	237°K
	10^{-2} molar	316°K	260°K

Since concentrations of 10^{-2} molar are within range of optical measurements, this means that it is worthwhile to look for the interaction between two dinucleoside phosphates. The values in the above table can only be considered a rough approximation since we have used the same value of ΔH over a temperature range of from 40 to 90 degrees.

It is interesting to compare the above calculations with those of Applequist and Damle for the interaction of oligo A's. They find that ApA should have a T_m of 210°K at 2×10^{-5} molar concentration.[3] This is in good agreement with the above values. If ΔH per residue is -7.5 kcal, Applequist and Damle predict that a change in concentration by a factor of 10 will shift T_m by 14 degrees. For a ΔH per residue of -4.4 kcal, the shift will be 26 degrees. These are smaller than our above values of about 20 and 40 degrees, respectively. If we pessimisticly assume an average shift in T_m of about 20 degrees per tenfold increase in concentration, this still means that formation of double strands by dinucleoside phosphates should be observable at temperatures around 0°C. We have chosen the interaction of ApA with poly U as a model for the above estimates, since this is expected to be one of the weakest possible interactions. Thus the above results are encouraging. Applequist and Damle have shown that increasing the chain length of

oligomers strongly stabilizes their interaction! Thus for chain lengths of 2, 3, 4, and 5 they find T_m's of -60°C, -25°C, 0°C, and +10°C, respectively, for interactions of A oligomers at less than 10^{-4} molar concentration.[3]

The one disturbing feature of the estimates discussed in this chapter is that they lead to the prediction that if dimers will interact to form double strands, the double strands will then, probably, aggregate. Since the tendency to aggregate should not increase with chain length, while the T_m certainly will, this is an argument for studying trimers or tetramers rather than dimers. But at the time this work was started, we had no complementary trinucleoside diphosphates at our disposal. Thus we decided to look for the interaction of complementary dinucleoside phosphates.

There are ten possible sets of dinucleoside phosphates which are complementary, if it is assumed that antiparallel interaction occurs in oligonucleotides. Four of these, GpC, CpG, ApU, and UpA, are self-complementary. The remaining 12 dinucleoside phosphates can form one pair each. We were unsure whether by ORD or UV spectra we could distinguish complementary interaction from just random stacking and other interactions which may not depend on sequence. Thus we decided to avoid studying self-complementary oligomers. Instead, we wanted to measure the optical properties of two oligomers separately, and then in a 1:1 mixture. Only if changes which occured in the latter were not observed in the pure components, would we be able to say that some kind of complementary interaction might be taking place. For this same reason we wanted to avoid the use of oligomers which contained 2 G's. These oli-

gomers are known to interact strongly with themselves (see Table II). Thus, if we saw optical changes as a function of temperature and concentration, we would not be sure whether the G's were interacting with themselves or their complements. Of course, if there were no spectral changes in the pure solution of GpG, we would be safe; but evidence presented before seems to indicate that interactions between G's are at least as strong as what we can expect for a G-C pair. But we wanted to have at least one G-C pair in our dimer complex, since evidence cited previously and theoretical estimates show that G-C pairs are more strongly bound than A-U.[158]

Consideration of the above conditions leads to the result that the following dimer interactions are desirable for study:

ApC with GpU	CpA with UpG
ApG with CpU	GpA with UpC

At the time our experiments were started, we had a dependable supply of all of the above dinucleoside phosphates except UpG and GpA. Thus our choice finally fell on the interaction of ApC with GpU.

In the future we plan to study the interaction of trinucleoside diphosphates with their complements. The constraints mentioned above suggest that it is preferable to study sets of compounds of the following types, where X is A or U and Y is complementary to X:

XpGpY with XpCpY	XpGpX with YpCpY
GpXpC with GpYpC	GpXpG with CpYpC
GpCpX with YpGpC	CpGpX with YpCpG

These arguments can easily be extended to provide suggestions for studying the interactions of higher oligomers. It seems to us advisable to avoid oligomers containing runs of G and to especially avoid oligomers which are self-complementary. Of course, if oligomers react to form parallel double strands, then no oligomer is self-complementary.

5. Dimer Interactions--Experimental Methods

(a) Solutions of Dinucleoside Phosphates

Lyophilized GpU was obtained from Dr. Warshaw. The sample used in these experiments was number 22D.[163] ApC was obtained from the California Biochemical Corp. (lot #45565). Subsequent work has shown that this lot of ApC contains a few percent of an impurity which is probably a mononucleoside. 0.01 M phosphate buffer was prepared, as described in Chapter II. The pH was 7.0 and NaCl was added to make the solution 0.5 molar in salt. This salt concentration was chosen to try to stabilize intermolecular interactions.[134] The dry dinucleoside phosphates were dissolved in buffer to make a concentration of about 2 mg per ml. This is about one half as concentrated as we would like. But preliminary work showed that GpU will not dissolve to the extent of 3.4 mg per ml at room temperature. The solutions were stored in the refrigerator for a few hours to make sure that no dinucleoside phosphate would precipitate when the temperature was lowered.

Aliquots of the above two solutions were diluted 1:100 using a microsyringe, and the absorption spectra were measured. Using extinction

coefficients at 260 mμ for ApC (1.05×10^4 per residue) and GpU (1.06×10^4 per residue) we were able to determine the concentration of the initial solution.[82] These concentrations were then adjusted by adding buffer to make the final concentrations of ApC 5.02×10^{-3} moles of residue per liter, and GpU 5.14×10^{-3} moles of residue per liter. A 1:1 mixture of ApC and GpU was prepared by mixing .400 ml of the ApC solution with .385 ml of the GpU solution. Thus the mixture of the two dinucleoside phosphates contained 2.56×10^{-3} molar (per residue) ApC and 2.52×10^{-3} molar (per residue) GpU. This and all of the following dilutions were performed with long tipped measuring pipettes.

Three serial dilutions of 1:10 with buffer were made from each of the above solutions. This resulted in solutions of approximately 5×10^{-3}, 5×10^{-4}, 5×10^{-5}, and 5×10^{-6} molar per residue of ApC, GpU, and a 1:1 mixture. The pipetting was checked by weighing, and the largest errors were of the order of one part in 300. Thus, these errors are ignored in the following treatment. The concentrations of oligomers in each of the solutions were computed from the known concentration of samples which gave O.D.'s of approximately .5 in a 1 cm cell. Thus an error in extinction coefficient would merely multiply all concentrations by a constant. Solutions were stored frozen in glass stoppered bottles until an hour before measurement. Then the samples were thawed and allowed to sit in an ice bath until we were ready to use them. In one experiment, solid $MgCl_2$ was added to a solution to bring the concentration of Mg^{++} up to 0.05 molar.

These experiments consumed a total of about 2 mg of each dinucleoside phosphate. (In our opinion, using standard cells, it would be impossible to use very much less material.) With ORD microcells, which we obtained after the conclusion of these experiments, it is now possible to perform an experiment like the above using only about 0.4 to 0.7 mg of each dinucleoside phosphate. Of course, the quantities needed are no problem in the case of dimers which are now commercially available. These comments are included for those who may wish to extend this work to higher oligomers.

(b) Calibration of Cells

All cells were cleaned with chromic acid cleaning solution and rinsed 10-20 times with distilled water before use. The relative path lengths of a series of cells were determined using a method described by Brode.[21] A standard solution of $KCrO_4$ in 0.05 M KOH was made to have an optical density of approximately 0.5 in a 0.1 mm (nominal) cell. Serial dilutions of this solution were made, each time diluting 10.0 ml to 100 ml. The spectrum of each dilution was measured in a cell of the appropriate path length, and Beers law was assumed to hold over the range of concentrations used. All spectra were run against air, and then a blank containing 0.05 M KOH was run in the same cell. The two readings at 373 mμ were subtracted, and the difference was used as an indication of path length. The results are shown on the following page; the path lengths given are relative to cell #A3.

Cell	Nominal Path Length	$O.D._{373}$-Blank	Path Length
A1	10. mm	.485	9.98 mm
A3	10. mm	.486	10.00 mm
B2	1.0 mm	.491	1.010 mm
B1	1.0 mm	.497	1.027 mm
X	0.1 mm	.610	0.1255mm
E2	100. mm	.461	98.8 mm

Note the good agreement between cell A1 and A3. This is an indication that it is not too foolhardy to neglect cell calibration for 1 cm cells. The very bad results for cell E2 are probably due to absorption of the sample onto the walls of the container. Almost all of our spectra run in the 10 cm cell have absorbances which are too low, and we have discarded all of these results. The path length of cell X, which we had borrowed from the Laboratory of Chemical Biodynamics, has been independently checked by other workers.[198] They reported a path length of 0.125 mm, using a method which does not assume Beers law. This result is in excellent agreement with the one reported above, and thus indicates that our path lengths reported relative to cell A3 are probably within 1% of the true path length.

(c) Optical Measurements

All optical measurements were performed as described in Chapter III, with the following exceptions. All UV spectra were measured against air and the solvent blank was run in the same cell and then subtracted point

by point. Measurements made at 2°C were done using a thermostated cell holder. The temperature of a cell containing water was measured with a thermocouple, and the cooling time was calibrated. Thereafter, no direct temperature measurements were made. At least two times the necessary cooling time was used in all cases. Reproducibility and time dependence were checked by measuring some spectra in duplicate after letting the solution sit at the temperature being used for about one hour. No time dependence was observed in any of our UV spectra on ORD curves. Glass stoppered or teflon stoppered cells were used in all cases. The cells used in the experiments reported in the next section are A3, B2, and X.

(d) Treatment of Optical Data

The optical data was treated as described in Chapter III. All results described in this and the following sections are expressed in units per nucleoside residue.

6. Dimer Interactions--Experimental Results

The ORD of ApC, GpU, and a 1:1 mixture at 2°C, is shown in Figures 1, 2 and 3. In each case the results at three concentrations are shown. These concentrations are nominally 5×10^{-3}, 5×10^{-4}, and 5×10^{-5} molar; the exact values were given in the last section. We had difficulty maintaining a constant base line using the 1 cm cell, since it is very difficult to place this cell reproducibly into the thermostated temperature block. Thus, base line corrections have been made with the 1 cm data. These consisted of adding the same constant factor to the curve

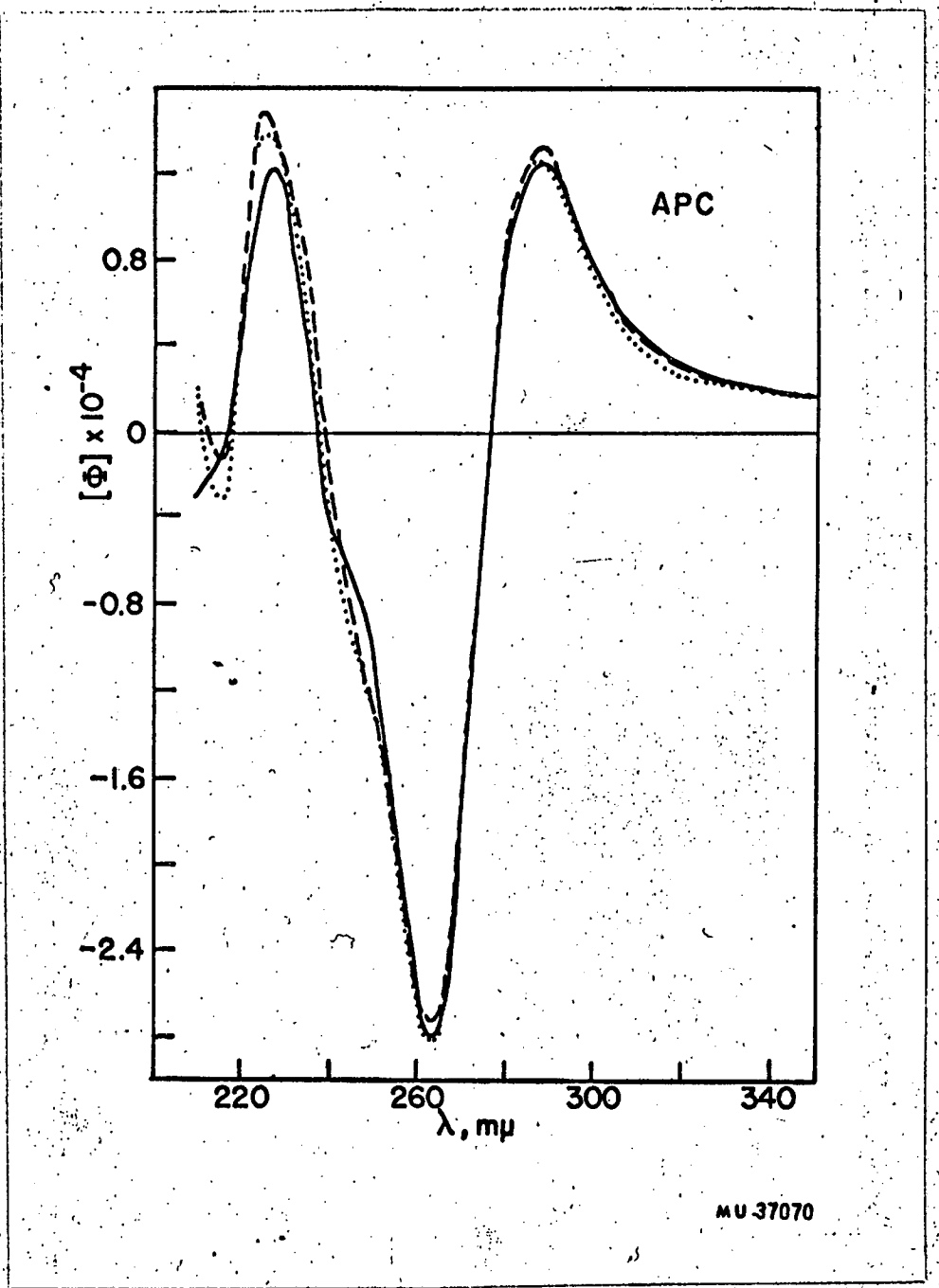


Figure 1. Optical rotatory dispersion of ApC at 2°C as a function of concentration.

——— 5×10^{-3} molar in nucleoside residues

– – – – 5×10^{-4} molar in nucleoside residues

. . . . . 5×10^{-5} molar in nucleoside residues

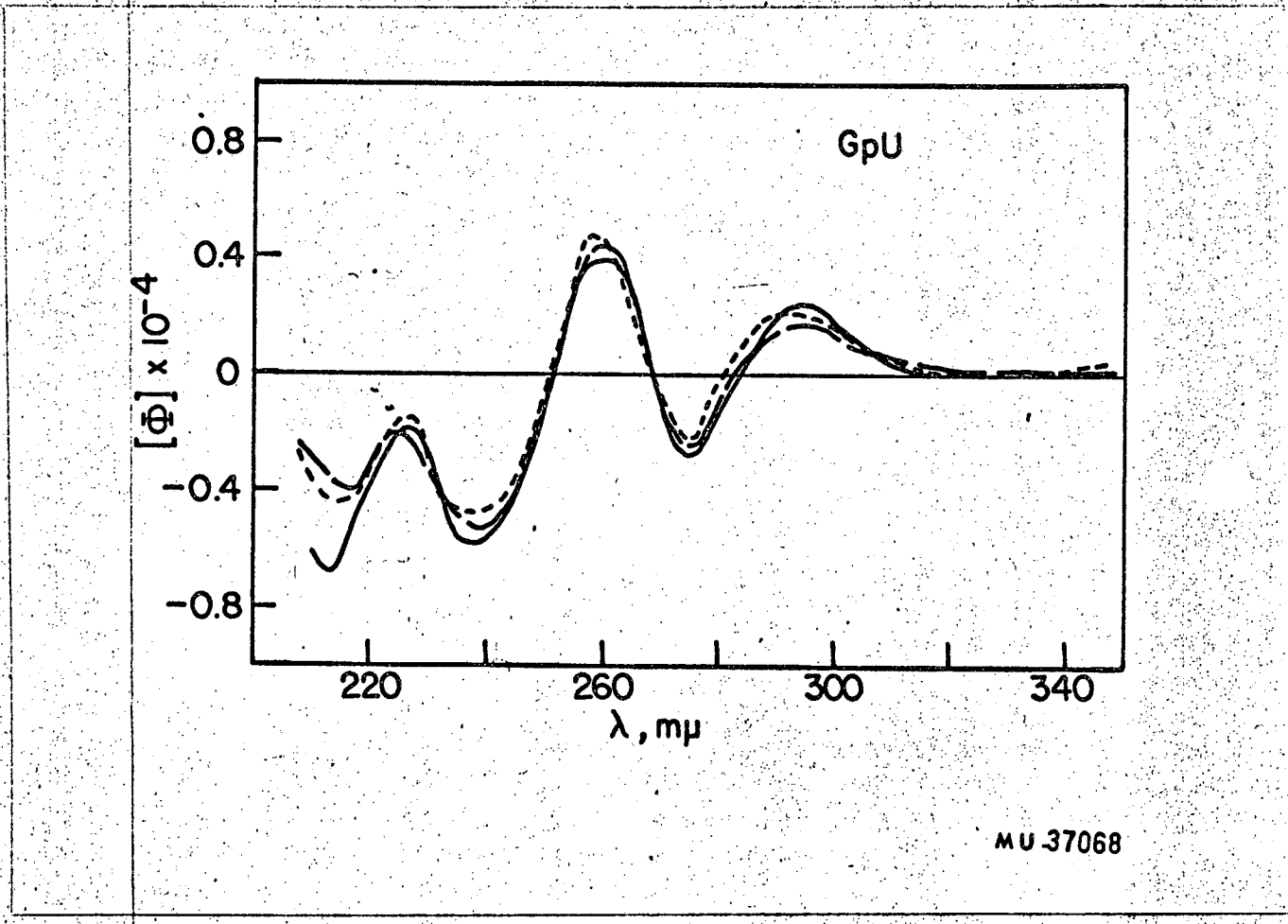


Figure 2. Optical rotatory dispersion of GpU at 2°C as a function of concentration.

——— 5×10^{-3} molar in nucleoside residues

— — — 5×10^{-4} molar in nucleoside residues

- - - - - 5×10^{-5} molar in nucleoside residues

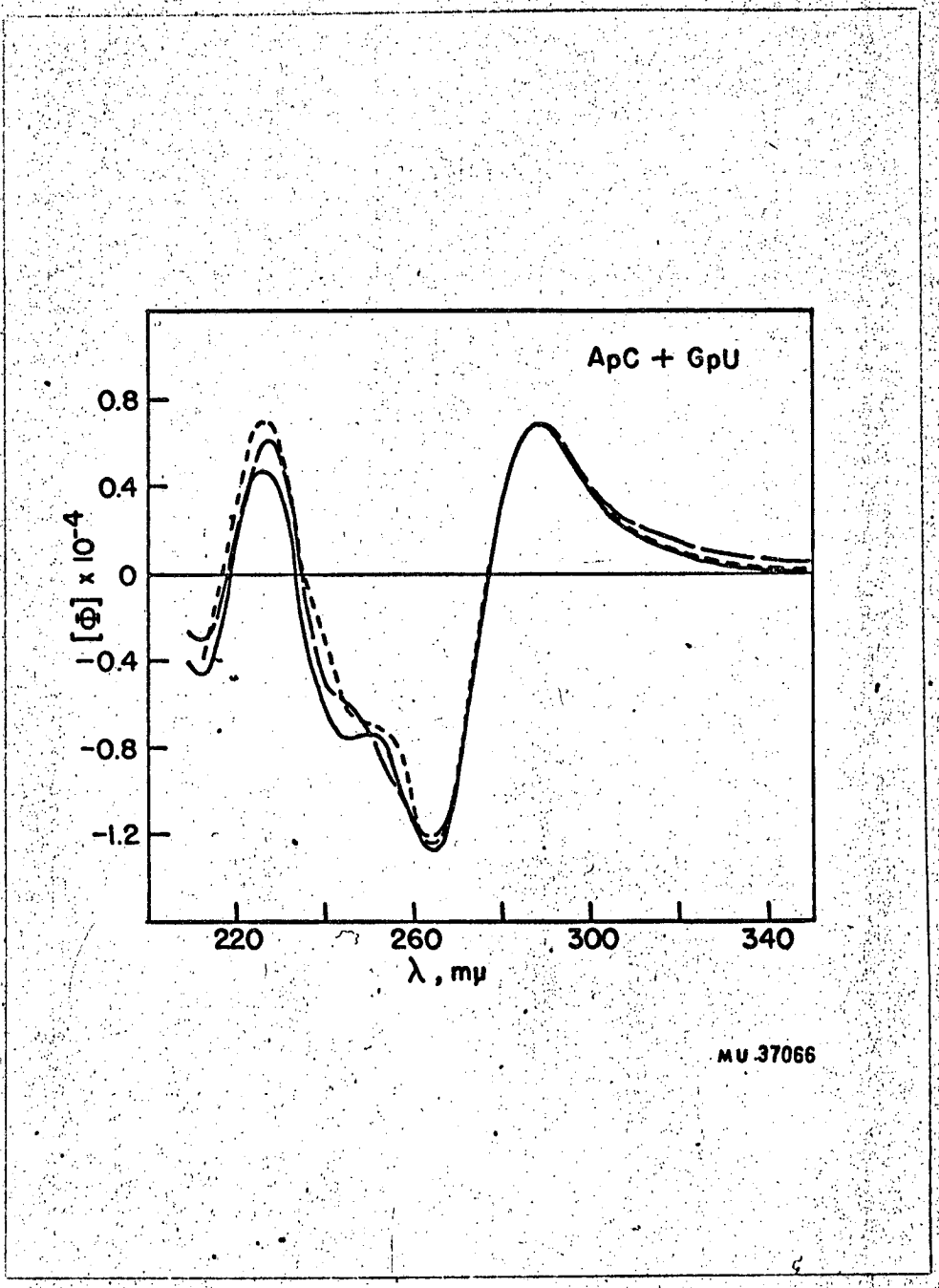


Figure 3. Optical rotatory dispersion of a 1:1 mixture of ApC and GpU as a function of concentration.

——— 5×10^{-3} molar in nucleoside residues

— — — 5×10^{-4} molar in nucleoside residues

- - - - - 5×10^{-5} molar in nucleoside residues

at every wavelength. This correction was chosen to make the curves taken in a 1 cm cell coincide at long wavelengths with the data obtained with shorter and longer path lengths. The correction factor is 2×10^{-3} molar rotation for GpU, and 4×10^{-3} molar rotation for the 1:1 mixture. We are sure that this correction is simply to overcome experimental artifacts, and does not mask any differences between the ORD at various concentrations.

Within the range of experimental error, the curves in Figures 1 through 3 show no systematic change as a function of concentration. The only possible exception to this are the discrepancies at very short wavelength among the three different concentrations of the 1:1 mixture. We feel that these results are negative; in fact, the excellent agreement between data at various concentrations could almost serve as a test of the reproducibility of our spectropolarimeter. Thus, if there are any intermolecular interactions between dinucleoside phosphates under our experimental conditions, these interactions do not affect the ORD. Of course, another interpretation would be that the dimers are aggregated throughout our concentration range. But if these dimers are completely aggregated at 10^{-4} molar concentration, then certainly GpC will also be an aggregate. But evidence cited in Chapter III shows that this is not the case for solutions of 10^{-4} molar GpC.

The UV spectra of ApC, GpU, and a 1:1 mixture at room temperature, are shown in Figures 4, 5 and 6. These are the same solutions whose ORD is shown in Figures 1 through 3. We have also measured the UV spectra of most of these solutions at 2°C. These results are not shown here graphically since they are very similar to the room temperature results.

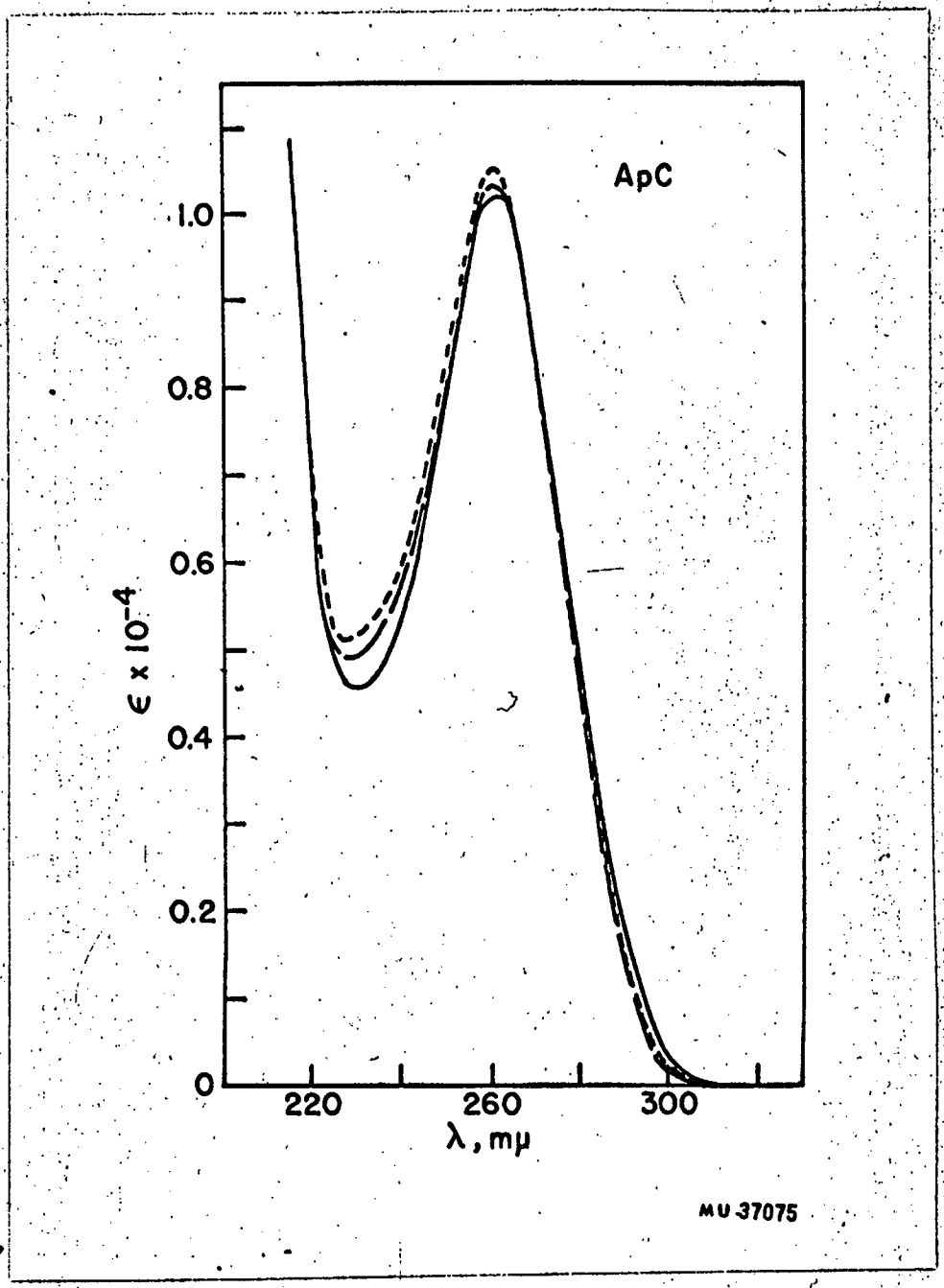


Figure 4. Absorption spectrum of ApC at room temperature as a function of concentration. Results are plotted as molar extinction per residue.

———— 5×10^{-3} molar in nucleoside concentration

— — — 5×10^{-4} molar in nucleoside concentration

- - - - - 5×10^{-5} molar in nucleoside concentration

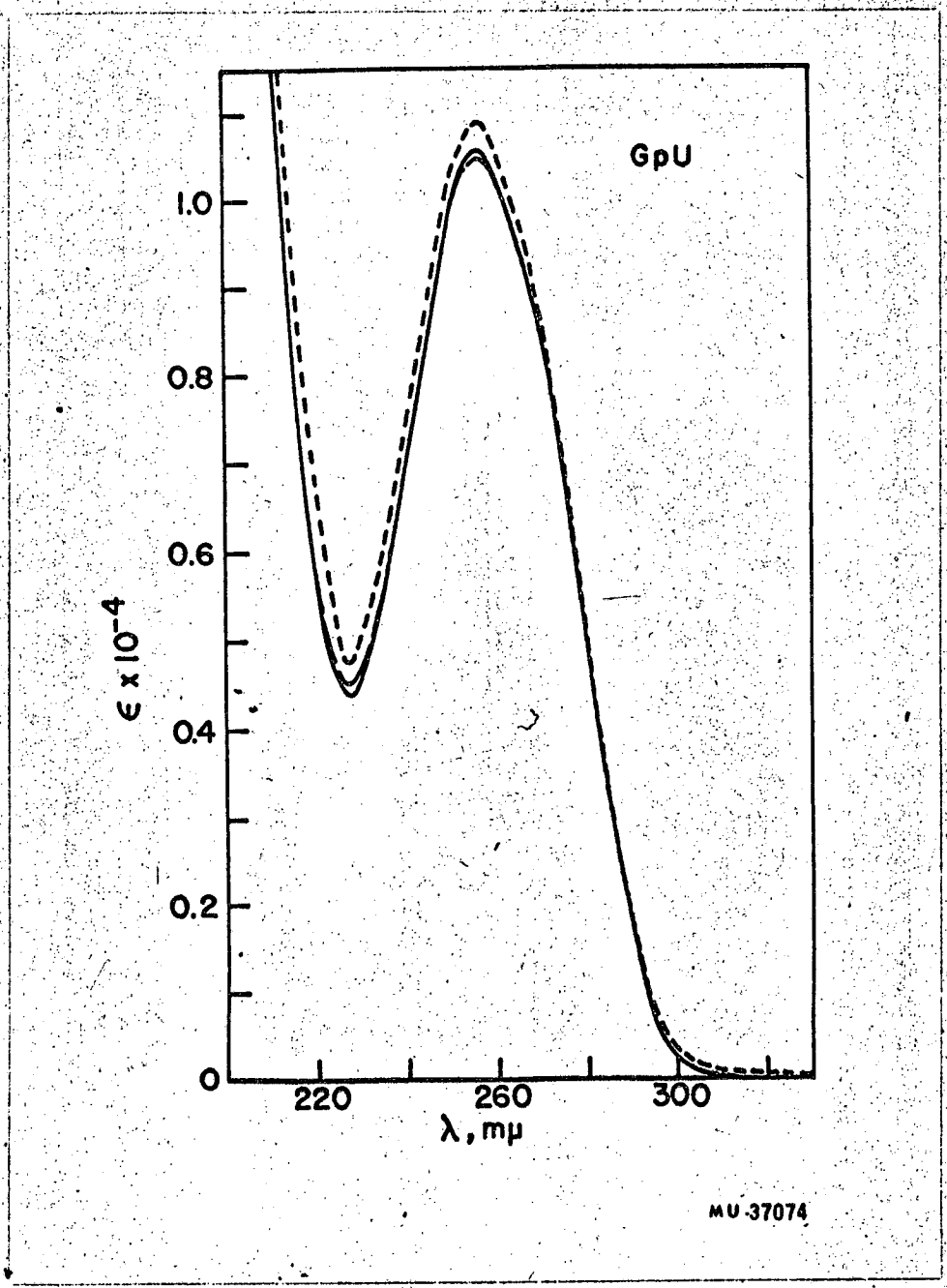


Figure 5. Absorption spectrum of GpU at room temperature as a function of concentration. Results are plotted as molar extinction per residue.

———— 5×10^{-3} molar in nucleoside concentration

— — — 5×10^{-4} molar in nucleoside concentration

- - - - - 5×10^{-5} molar in nucleoside concentration

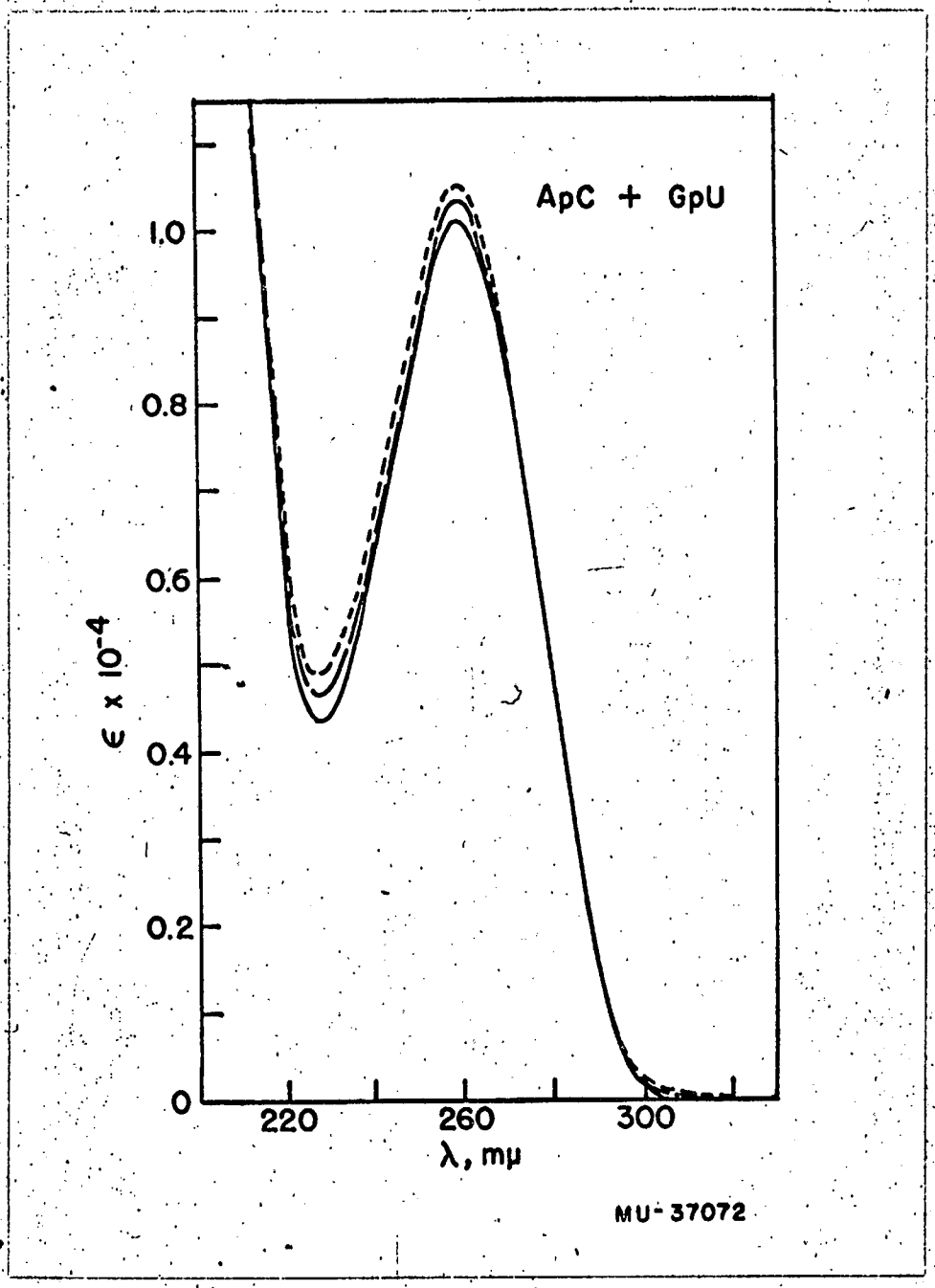


Figure 6. Absorption spectrum of a 1:1 mixture of ApC and GpU at room temperature as a function of concentration. Results are plotted as molar extinction per residue.

———— 5×10^{-3} molar in nucleoside concentration

— — — 5×10^{-4} molar in nucleoside concentration

- - - - - 5×10^{-5} molar in nucleoside concentration

Values at selected wavelengths can be found in Table III for spectra taken at both temperatures. The signal to noise is much higher for UV spectra than ORD, and it is possible to observe a small temperature dependence for the spectra of ApC, GpU, and a 1:1 mixture. If there had been a corresponding change in the ORD of this order of magnitude, we would not have been able to see it. The largest change in extinction at the absorption peak occurs in ApC + GpU (1:1), where a 4% decrease is observed as the concentration is increased from 5×10^{-5} to 5×10^{-3} molar. The percentage change is larger at smaller wavelength, but that region of the spectrum is often very sensitive to impurities. One would have to be cautious about making an estimate of the order of magnitude of this effect from data at 220 or 230 mμ. Over the same concentration range, the maximum extinction of ApC changes by 3-1/2%, and the change in GpU is about 3%. In each case, the absorbance decreases as the concentration increases. There is also some evidence for an increase in extinction at long wavelengths for GpU and the mixture as the concentration is raised. The effect in ApC is the opposite, but is sufficiently small to be explainable by a small error in setting the starting wavelength of the scan. It should be mentioned that although these concentration effects are very small they are larger than our volumetric error.

Since the extinction coefficients we observe are good to one percent, we can calculate the expected spectrum of a mixture of two dinucleoside phosphates to that degree of accuracy. To do this, we shall assume that the optical properties do not change more than 1% over the two fold range of concentration between our spectra of pure dimer and dimer in a 1:1 mixture. The equation used is

$$\varepsilon_{mixt}(\lambda) = 0.4973\ \varepsilon_{ApC}(\lambda) + 0.5026\ \varepsilon_{GpU}(\lambda) \qquad (7)$$

TABLE IIIa

Optical Parameters for ApC, GpU, and a 1:1 Mixture

Optical Rotatory Dispersion

Substance	Cell	Molar Rotation x 10^{-4}			Long Wavelength (mμ)			Short Wavelength (mμ)	
		$[\phi]_p$	$[\phi_t]$	$\lvert[\phi_p]-[\phi_t]\rvert$	λ_p	λ_o	λ_t	λ_o	λ_p
ApC	0.1 mm	1.22	-2.78	4.00	287.5	277	264	237.5	227.5
ApC	1.0 mm	1.33	-2.70	4.03	287.5	276	262.5	237	225
ApC	10. mm	1.22	-2.83	4.05	286	276.5	262.5	237.5	225
ApC	100. mm††	1.19	-2.92	4.11	285	277	262.5	237	225
GpU	0.1 mm	.39	-.24	.63	293	283.5	275	269	260
GpU	1.0 mm	.44	-.23	.67	292.5	282	275	268	260
GpU	10. mm†	.48	-.21	.69	292.5	282.5	275	267.5	258.5
GpU	100. mm††	.25	-.37	.62	295	285	275	267	258.5
ApC+GpU	0.1 mm	.67	-1.26	1.93	292	277	265	234	227.5
ApC+GpU	1.0 mm	.69	-1.22	1.91	287.5	277	264	234.5	227.5
ApC+GpU	10. mm†	.69	-1.25	1.94	287.5	277	265	235	227.5
ApC+GpU	100. mm††	.55	-1.38	1.93	286	278.5	264	234	227.5
Calculated from spectra taken separately*									
ApC+GpU	0.1 mm	.70	-1.26	1.96	287.5	277.5	264	232	226
ApC+GpU	1.0 mm	.74	-1.22	1.96	287.5	277	264	236	225
ApC+GpU	10. mm	.79	-1.20	1.99	287.5	276	265	235	226
ApC+GpU	100. mm††	.59	-1.37	1.96	285	278	264	232	225

†The baseline has been corrected for spectra taken of GpU and the 1:1 mixture in the 10 mm cell by subtracting .20 x 10^{-4} and .40 x 10^{-4} molar rotation, respectively.

††The 100 mm data is shown for illustrative purposes only. We cannot trust it, for reasons explained in the text.

*Calculated using the equation: $[\phi]_{mixt} = 0.4973[\phi]_{ApC} + 0.5026[\phi]_{GpU}$.

Note: t, p, and o refer to trough, peak, and cross-over, respectively.

TABLE IIIb

Absorption Spectra

Substance	Cell	Residue Extinction x 10^{-4} Room Temperature 280	260	230	Low Temperature 280	260	230
ApC	0.1 mm	.488	1.013	.458			
ApC	1.0 mm	.449	1.030	.494	.439	1.013	.485
ApC	10. mm	.460	1.049	.517	.444	1.025	.503
GpU	0.1 mm	.512	1.025	.454			
GpU	1.0 mm	.514	1.027	.462	.487	1.007	.468
GpU	10. mm	.533	1.058	.506	.512	1.039	.490
ApC+GpU	0.1 mm	.480	1.010	.444			
ApC+GpU	1.0 mm	.482	1.032	.474	.458	1.002	.462
ApC+GpU	10. mm	.491	1.046	.500	.473	1.026	.485
ApC+CpU	1.0 mm**				.470	1.008	.464
Calculated from spectra taken separately††							
ApC+GpU	0.1 mm	.500	1.019	.456			
ApC+GpU	1.0 mm	.482	1.028	.478	.463	1.010	.476
ApC+GpU	10. mm	.497	1.054	.512	.478	1.031	.496

**In the presence of 0.05 M Mg^{++}.

††Calculated using the equation: $\epsilon_{mixt} = 0.4973\ \epsilon_{ApC} + 0.5026\ \epsilon_{GpU}$

Exactly the same equation is used for ORD, where extinction per residue is replaced by molar rotation per residue. This equation takes into account the slight difference in concentration between ApC and GpU in the mixture. The calculated and experimental ORD curves for the nominally 1:1 mixture are shown in Figures 7 and 8 for data taken in cells of 0.1 and 1.0 mm, respectively. The agreement of the two ORD curves at concentrations of 5×10^{-4} molar is within experimental error. The agreement between calculated and experimental mixtures at 5×10^{-3} molar is not as good. Differences at long wavelength and in the region from 240 to 250 mμ fall outside the normal range of experimental error. But we think it would be unwise to attach any great significance to these differences. The long wavelength differences could easily be explained by a small base line shift. And the discrepancy at shorter wavelength reappears for data taken at lower concentrations than are shown here. Numerical values for the amplitude ($|[\phi_P]-[\phi_T]|$) of the long wavelength Cotton effect of the 1:1 mixture of ApC and GpU are shown in Table III along with values calculated for a 1:1 mixture. The agreement between calculation and experiment for this quantity, which is base line dependent, is excellent. Using the above equation, we have also calculated the UV spectra expected for a 1:1 mixture of ApC and GpU. These results are not shown graphically because agreement between experiment and calculation is so good that most of the small differences observed would not be visible on a graph. From these results it may be concluded that, although some concentration dependence of UV spectra is observed, this cannot be attributed to a selective interaction between ApC and GpU. More likely, it represents some kind of random aggregation. The question

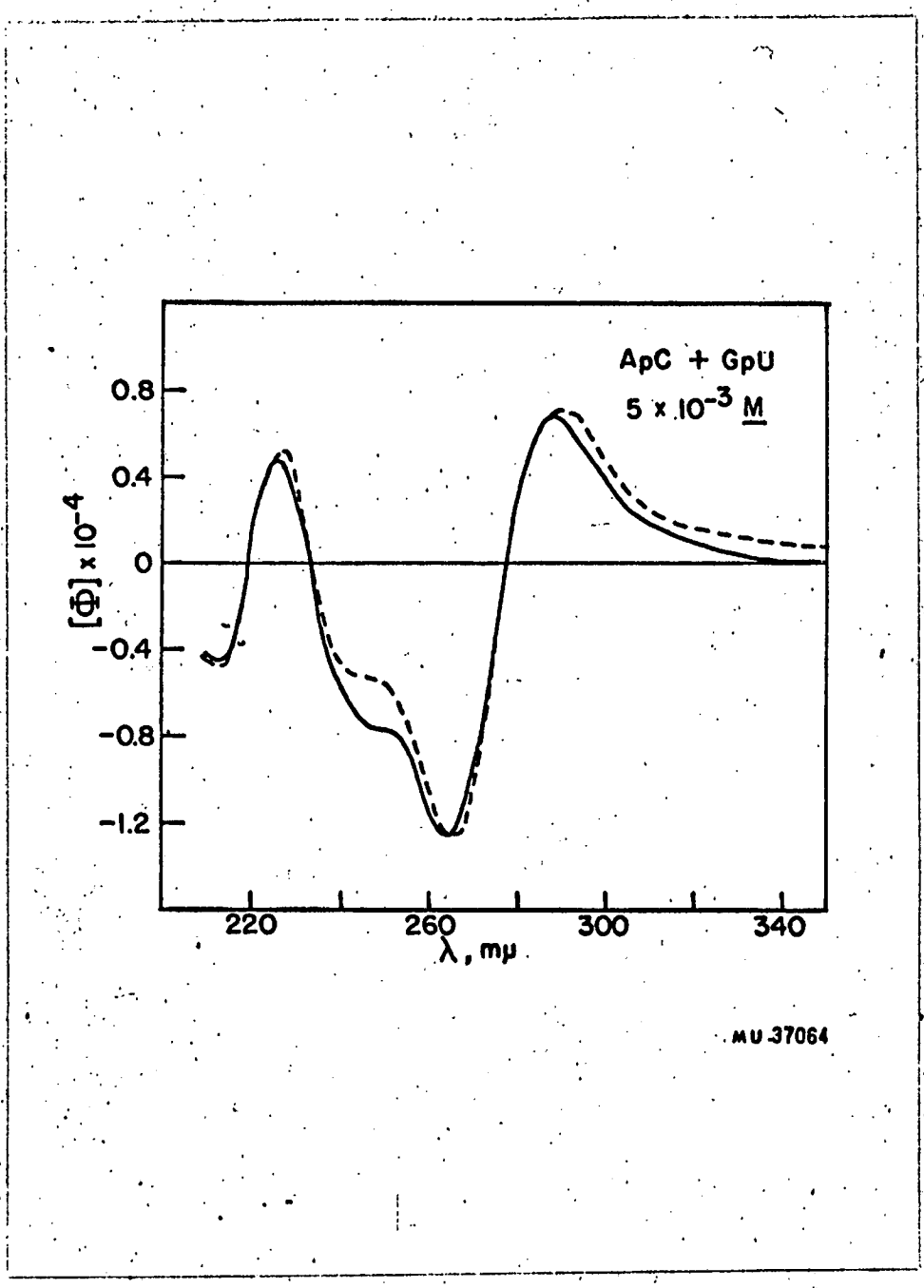


Figure 7. Optical rotatory dispersion of a 1:1 mixture of ApC and GpU. The concentration of total nucleoside residues is 5×10^{-3} molar.

———— experimental

– – – – – sum of ApC and GpU measured separately

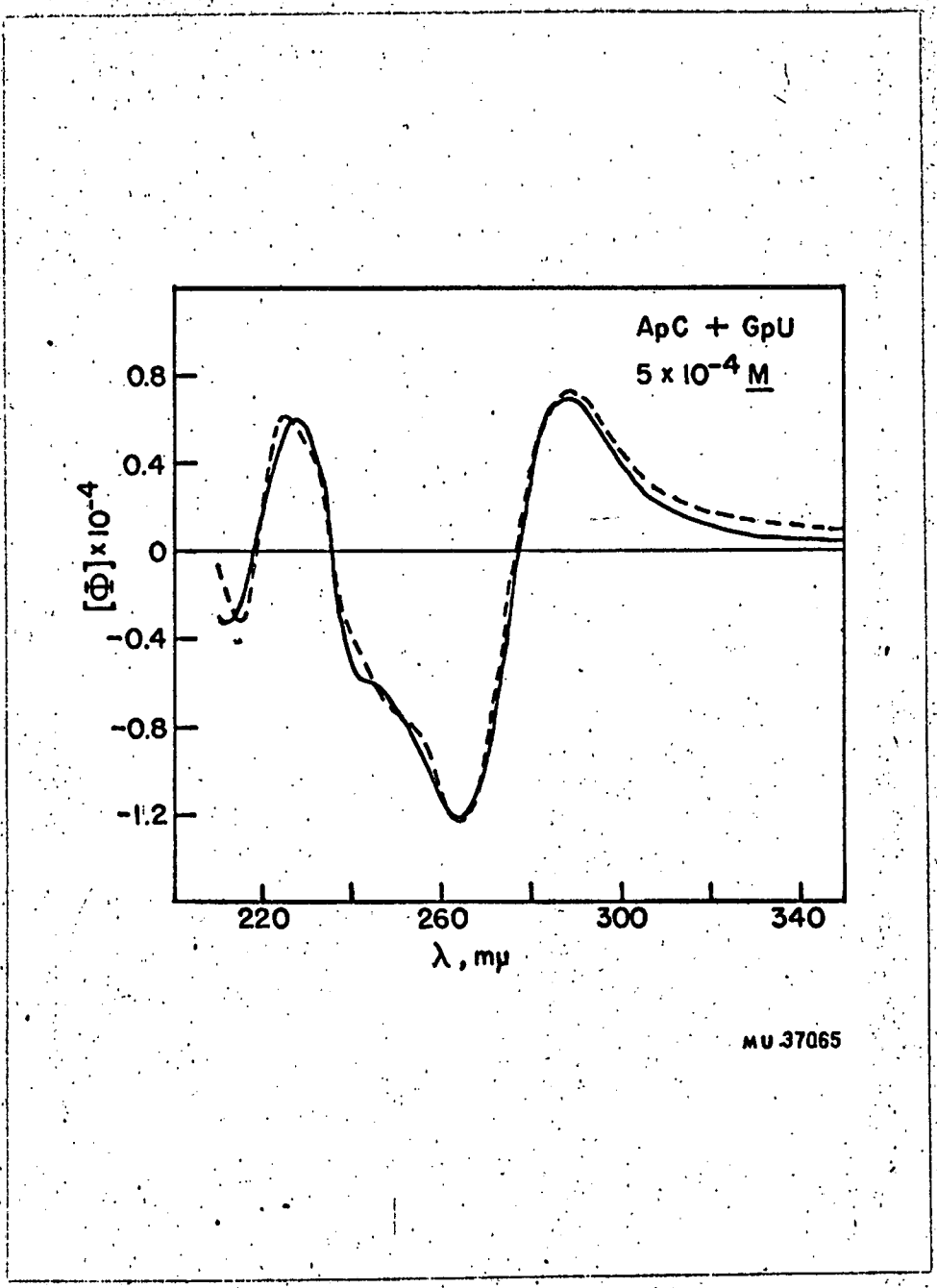


Figure 8. Optical rotatory dispersion of a 1:1 mixture of ApC and GpU. The concentration of total nucleoside residues is 5×10^{-4} molar.

——— experimental

– – – – – sum of ApC and GpU measured separately

which remains unanswered is why this aggregation does not cause measurable changes in the ORD. We can only conclude that the aggregation might lead to interactions whose contributions to the ORD cancel. If, for example, vertically stacked dimer interactions occur, with no preferential orientation of one dimer above the next, the changes in the ORD could be very small. In this case, there would still be a hypochromic effect.

In a further attempt to promote aggregation or any other interaction of dinucleoside phosphates, we added Mg^{++} ion to a sample of 5×10^{-4} molar 1:1 mixture and reran the absorption curve. We had originally avoided the use of Mg^{++}, since it is known to favor triple strand interactions, and we wanted to study double strands. But the addition of Mg^{++} produced absolutely no change in the absorption spectrum. Thus we have yet one more negative result.

Faced with such disappointing results, we feel entitled to ask the question: if base pairing or some kind of specific vertical stacking aggregation had occured, could we have seen it under our experimental conditions? Suppose that dimers aggregate to form tetramers, in which the aggregate resembled the conformation the dimers would have if they were joined by a phosphodiester bond. This may not be the most realistic model for a dimer aggregate, but we have chosen it because it is very easy to predict the ORD of such a conformation. The reaction we shall consider corresponds to the reaction

pure ApC: ApC + ApC = ApCpApC

pure GpU: GpU + GpU = GpUpGpU

1:1 mixt.: ApC + GpU = 1/4 ApCpApC + 1/4 GpUpGpU + 1/4 ApCpGpU + 1/4 GpUpApC.

The ORD of each of the products of the above reactions was calculated using the methods discussed in Chapter III. Thus, for example, for ApCpApC,

$$[\phi] = 1/4\,(4[\phi]_{ApC} + 2[\phi]_{CpA}) \quad . \qquad (9)$$

Unfortunately, we need to know the molar rotation of each of the dimers involved in the above calculation at 2°C. But we have data only for ApC and GpU at this temperature. So we have used 2°C molar rotation for these two dimers, and room temperature data for the other four which are needed. But since each of the dimers which appear when the aggregate is formed contributes only about 1/3 of the rotation, the errors are small. The ORD calculated for stacked dimer aggregates (using data at 5×10^{-5} molar) is compared with the experimental data at 5×10^{-3} molar concentration in Figures 9, 10 and 11. While the differences between calculated and experimental curves are still small, the differences are clearly outside experimental error. They are much larger than any of the differences among ORD taken at different concentrations. Thus, we can conclude that if 50% or more of the dimers had aggregated in the way we described above, we should have observed this.

A more reasonable guess for the conformation of dinucleoside phosphate aggregates would be to assume that all of the possible ways in which two dimers can stack are equally probable. Thus ApC would form aggregates of the type ApC-ApC, ApC-CpA, CpA-ApC, and CpA-CpA. If each of these aggregates preferred the same conformation as the corresponding tetramer we would again be able to calculate the ORD change expected for this process. But qualitatively, this should be similar to the simpler model we have already treated. The average result will still be to have

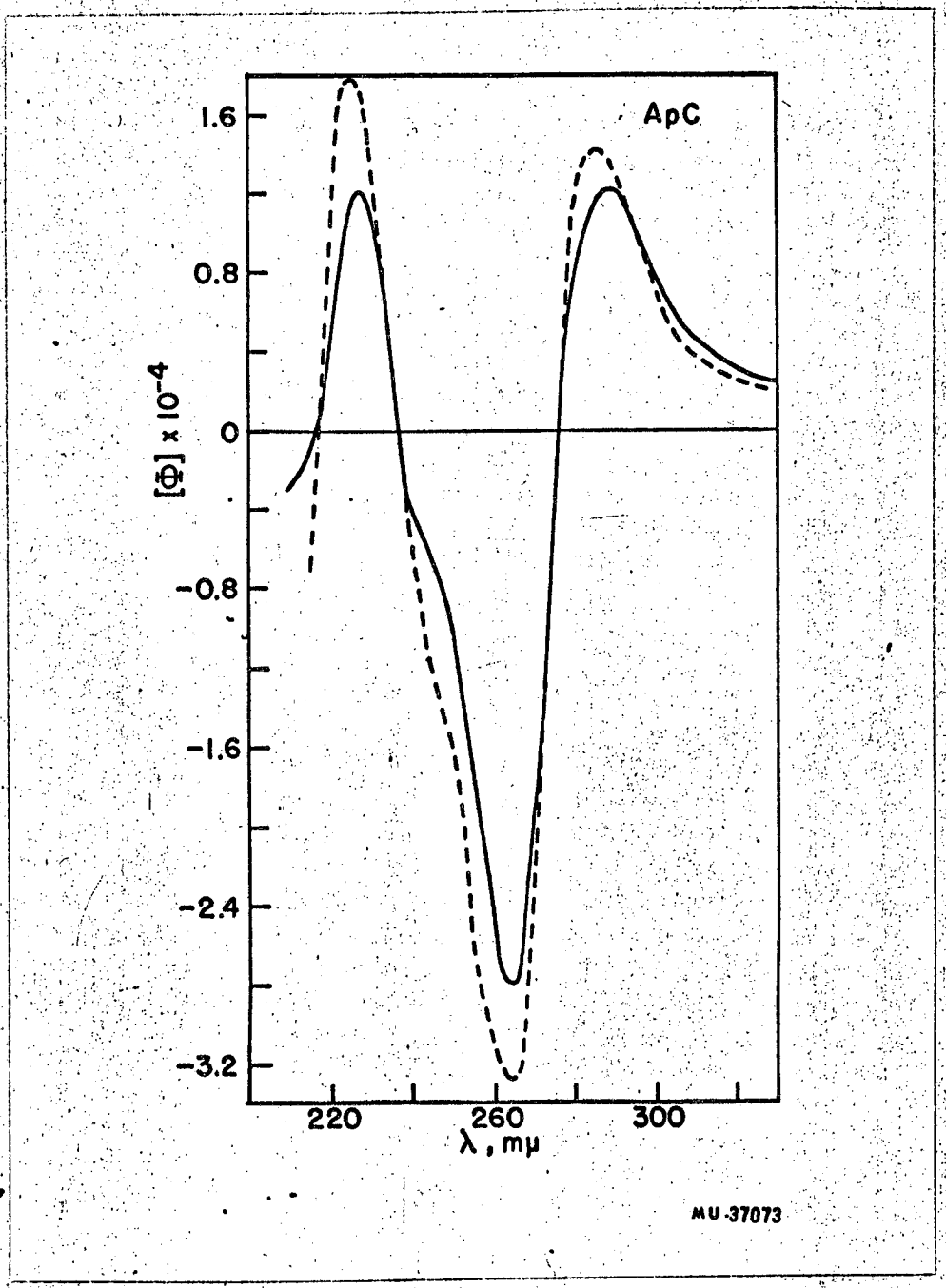


Figure 9. A possible effect of aggregation of ApC on the optical rotatory dispersion. See text for details.

———— experimental ORD at 5×10^{-3} molar residues

– – – – – calculated ORD of a stacked aggregate

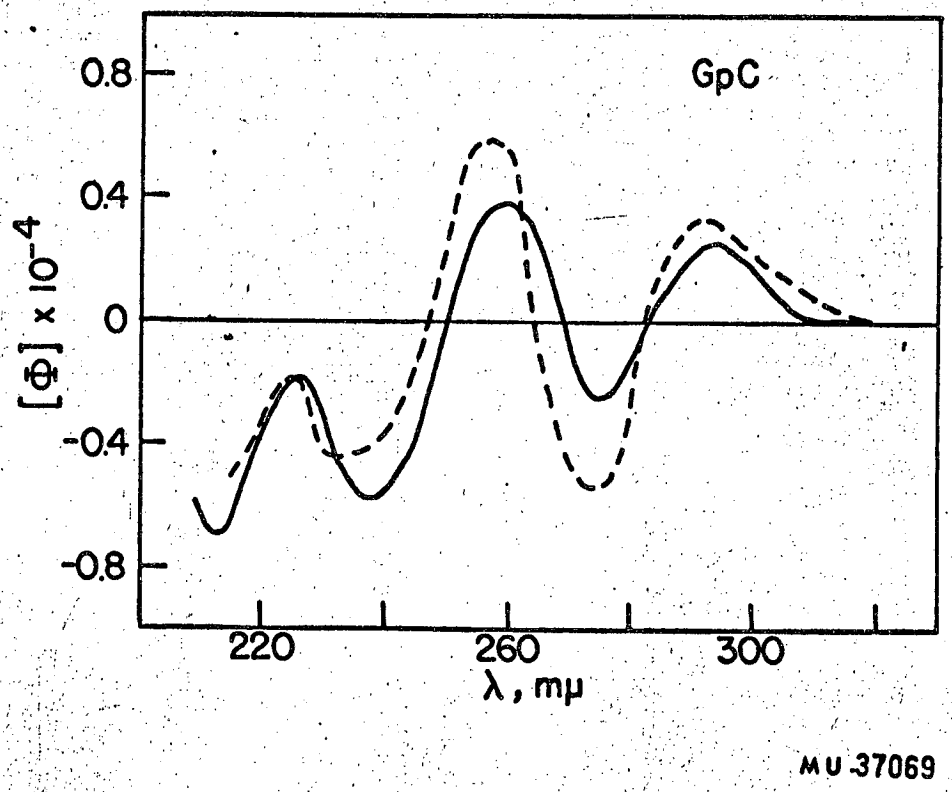


Figure 10. A possible effect of aggregation of GpU on the optical rotatory dispersion. See text for details.

_____ experimental ORD at 5 x 10^{-3} molar residues

_ _ _ _ _ calculated ORD of a stacked aggregate

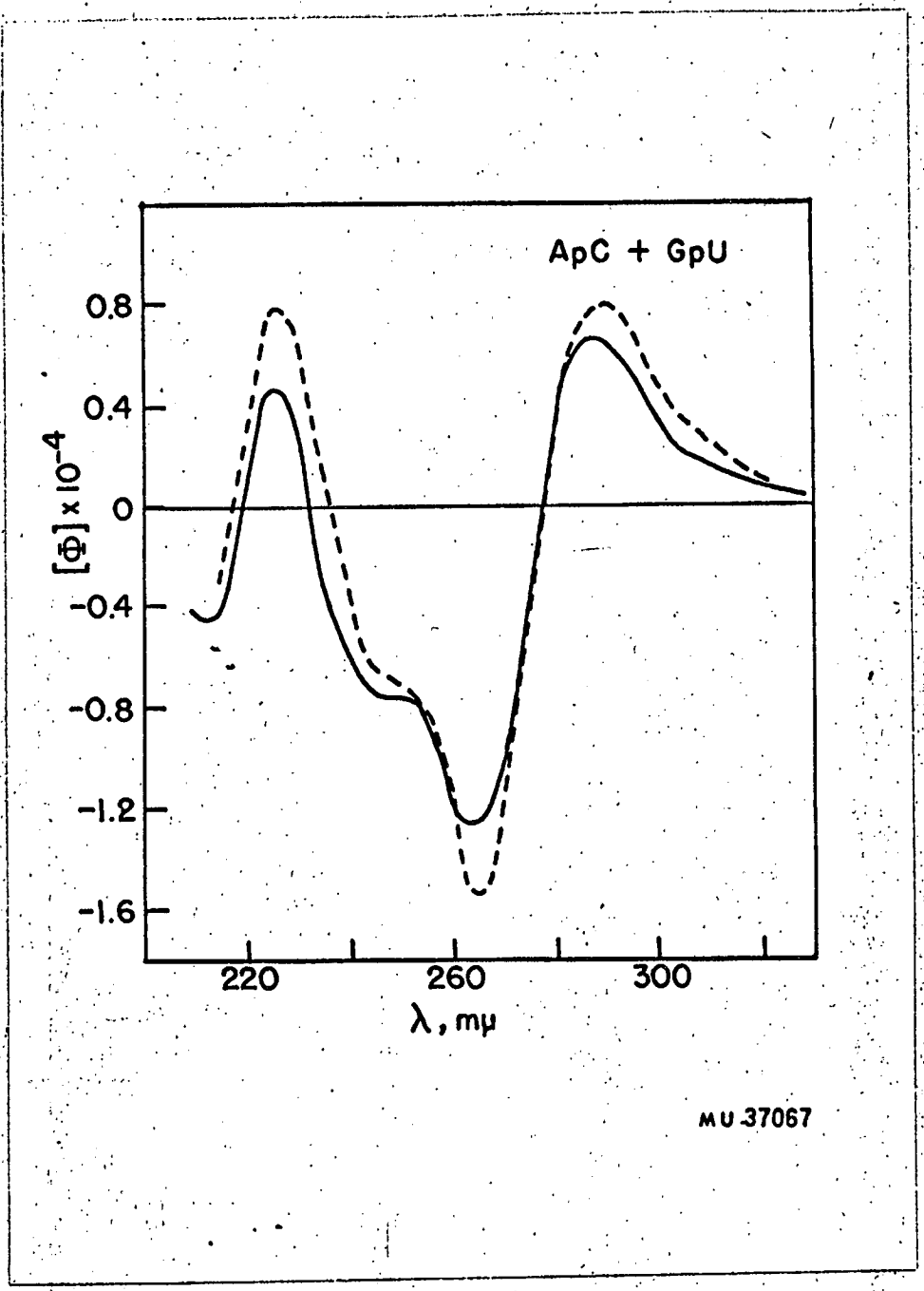


Figure 11. A possible effect of aggregation of a 1:1 mixture of ApC and GpU on the optical rotatory dispersion. See text for details.

——— experimental ORD at 5 x 10^{-3} molar residues

– – – – calculated ORD of a stacked aggregate

3/2 as much nearest neighbor interaction in the aggregate than in the dimer. But it is our feeling that the most logical guess for aggregate conformations would be to say that they are very random. In this case we cannot calculate the ORD of the aggregate, but it would be expected to be almost the same as the unaggregated dimers.

Suppose ApC and GpU had really interacted to form a double stranded dimer. Would we have been able to observe this by measuring the ORD? We cannot answer this question with as much confidence as our guesses for vertical aggregates. But methods which will be described in later sections permit us to make a rough estimate of the change in ORD upon forming an A-U or G-C pair. Since the double strand complex of ApC and GpU would have one of each, the ORD change expected can be approximated by the average for an A-U and a G-C pair. We cannot jump ahead to explain how this can be estimated, but we would like to include the result here. Figure 12 shows the ORD expected for double strand ApC/UpG. It is easily seen that if double strand formation had occured to any appreciable extent we should have seen it. Thus we must conclude that, in our hands, ApC and GpU do not interact to form a double strand.

7. ORD of Single Strand Homopolymers

Up till now we have limited our discussion to the optical properties of oligonucleotides. In this and the succeeding sections we shall extend the methods explained in Chapter III to polymers. In this section we will treat the ORD of single stranded polymers. The complications which ensue when double strands can be formed will be considered in later sections.

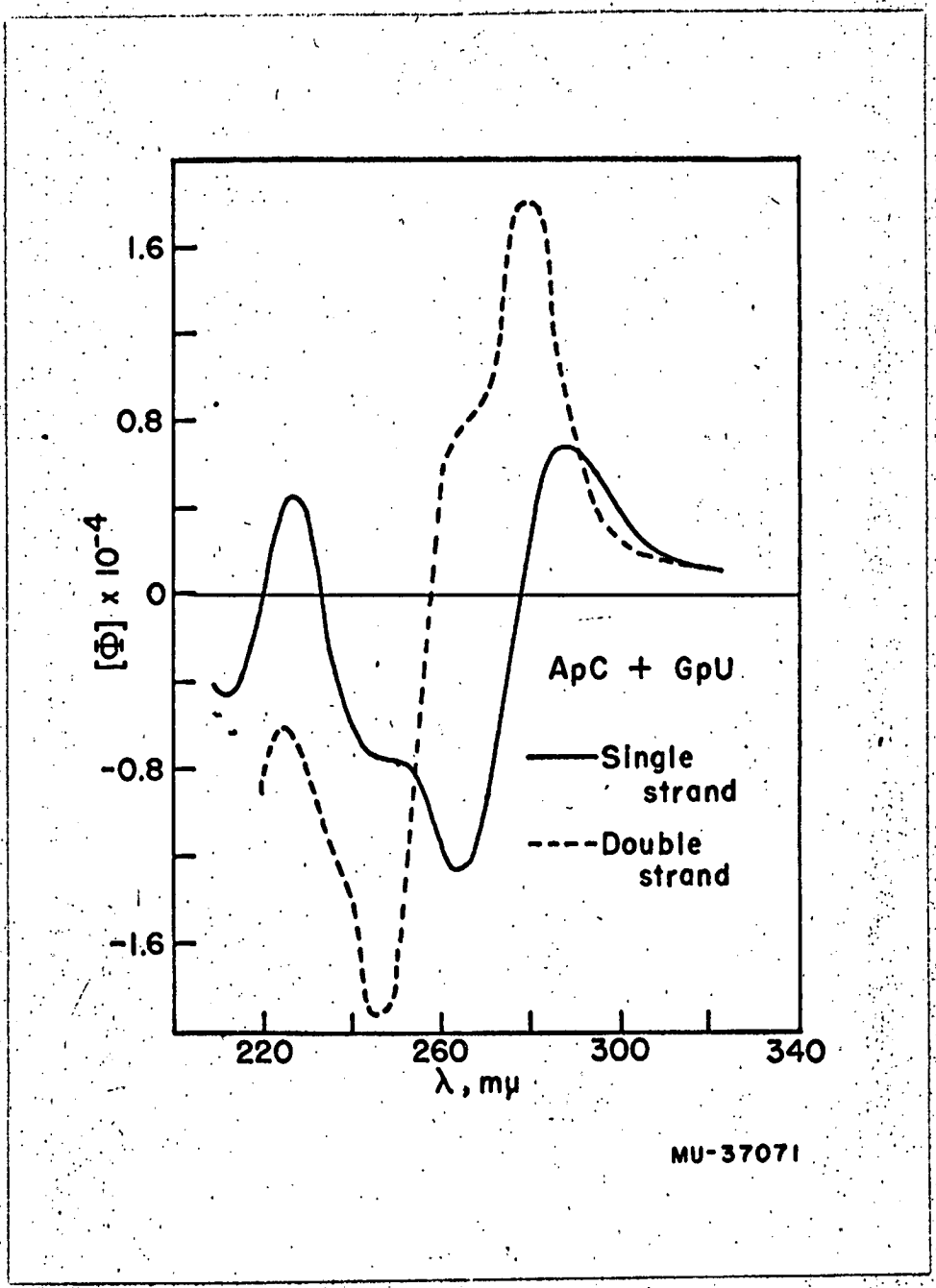


Figure 12. A possible effect of the formation of a specific double strand complex of ApC and GpU on the ORD.

——— experimental ORD of a 1:1 mixture at 5×10^{-3} molar residue

– – – – calculated ORD of ApC/GpU anti-parallel double strand

Evidence cited in Chapter III leads to the conclusion that poly A, poly U and poly C are single strands at pH 7 and room temperature. Thus if our nearest neighbor calculations are to hold any promise for predicting the ORD of RNA's they should give a reasonable approximation to the ORD of a homopolymer. To calculate the ORD of a homopolymer, equation (12) of Chapter III reduces to the simple form

$$[\phi]_{\text{poly } N}(\lambda) = 2[\phi]_{NpN}(\lambda) - [\phi]_{N}(\lambda) . \qquad (10)$$

We have used this equation and the dinucleoside phosphate data and monomer data described in Chapter III to calculate the ORD of poly A, poly C, poly U, and poly G. These results are shown in Figures 13, 14, 15, and 16, respectively. The experimental curve of poly A was taken from the data of Holcomb and Tinoco.[95] It was run at pH 6.9 and 0.15 M KCl, conditions very similar to those used for the dinucleoside phosphates. The poly U experimental curve had been digitized from the data of Sarkar and Yang.[195,251] The experimental conditions used were 0.15 M KF, pH 7.5. We think that these are still close enough to the conditions of our dimer data to permit a fair comparison. The data for poly C were obtained from Warshaw.[251] They are from a commercial sample of poly C (Miles Laboratory) run under the same conditions used for most of our dinucleoside phosphates. There are considerable discrepancies among the ORD results reported for poly C by various workers.[54,196] However, data obtained in our laboratory for poly C prepared by Chamberlin are in excellent agreement with the ORD shown in Figure 14.[135] Thus, we feel that differences among published experimental ORD curves for poly C primarily represent different choices of extinction coefficients. The value used for the results shown in Figure 14 is $\epsilon_{268} = 0.622 \times 10^4$.

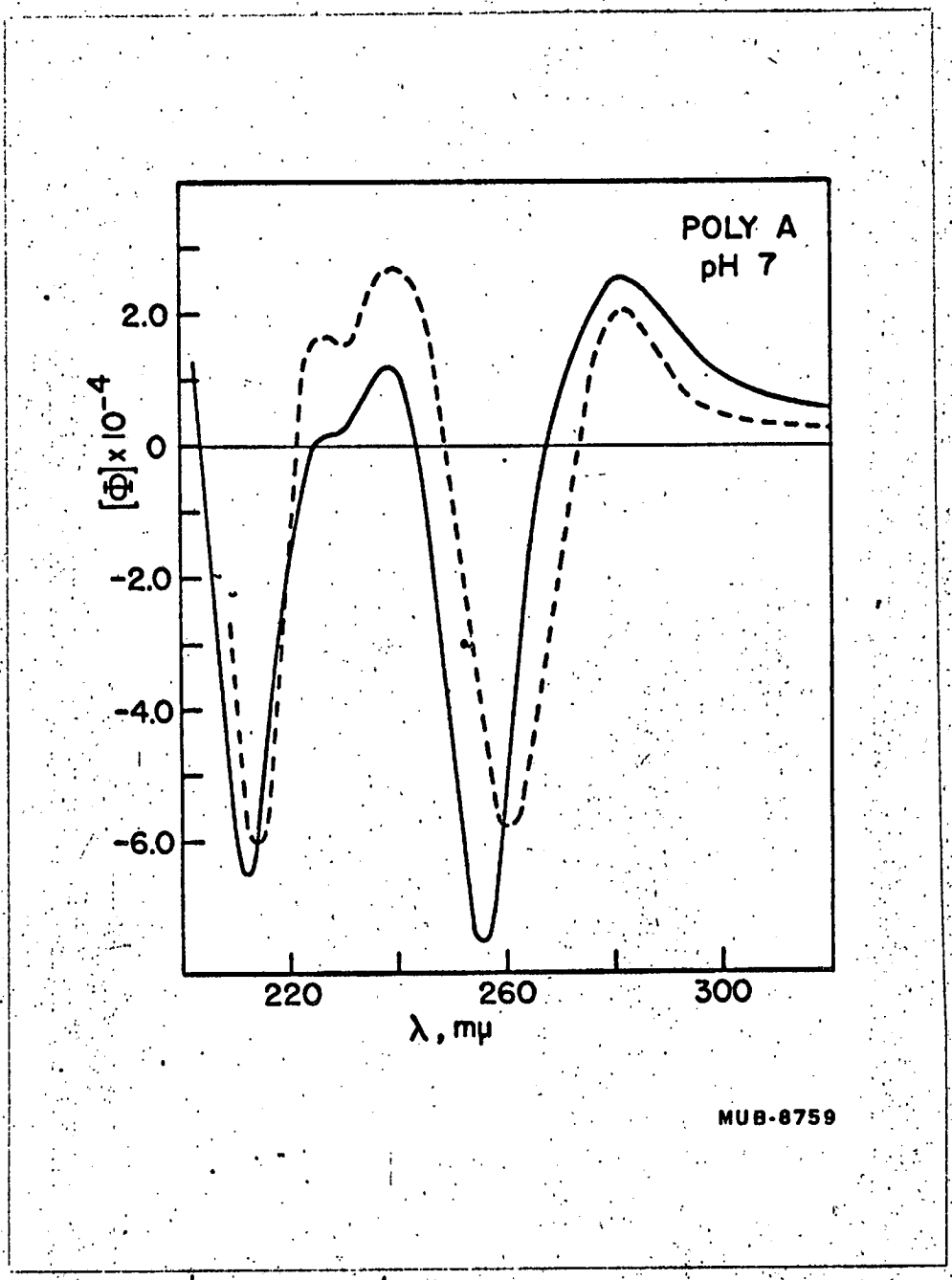


Figure 13. The optical rotatory dispersion of poly A at neutral pH.

——— experimental[95]

– – – – – nearest neighbor calculation

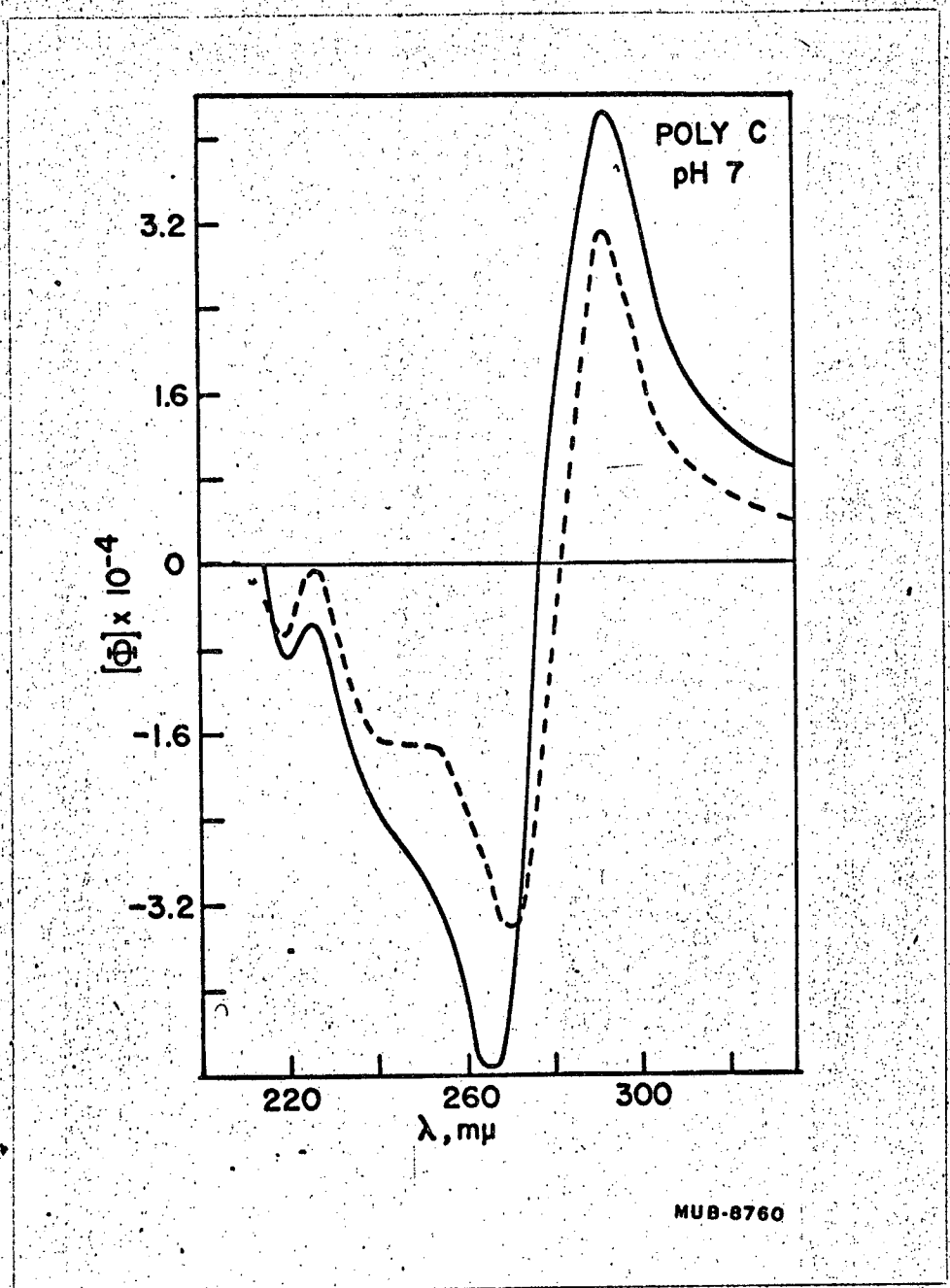


Figure 14. The optical rotatory dispersion of poly C at neutral pH.

——— experimental[251]

– – – – – nearest neighbor calculation

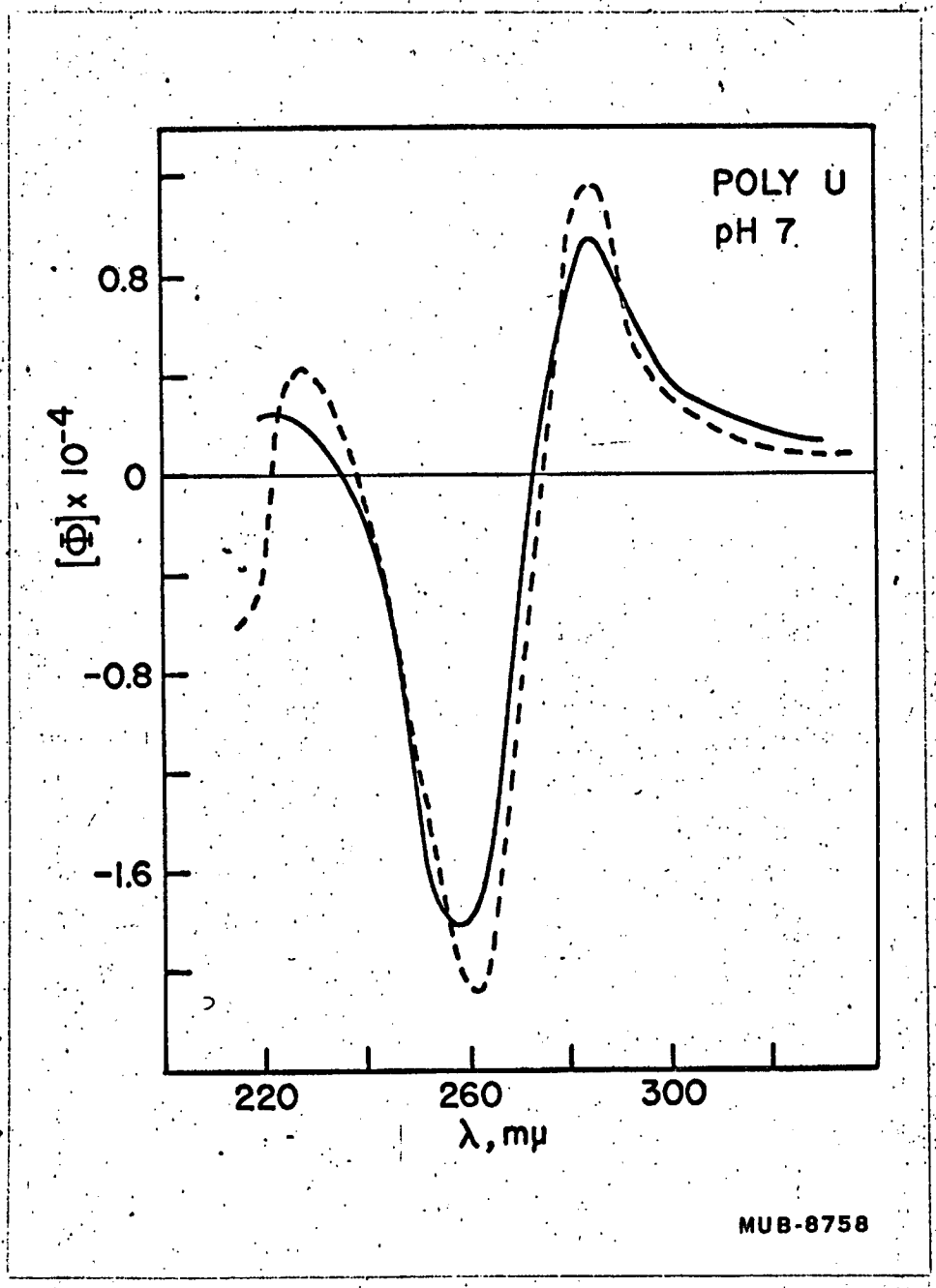


Figure 15. The optical rotatory dispersion of poly U at neutral pH.

——— experimental [105]

– – – – – nearest neighbor calculation

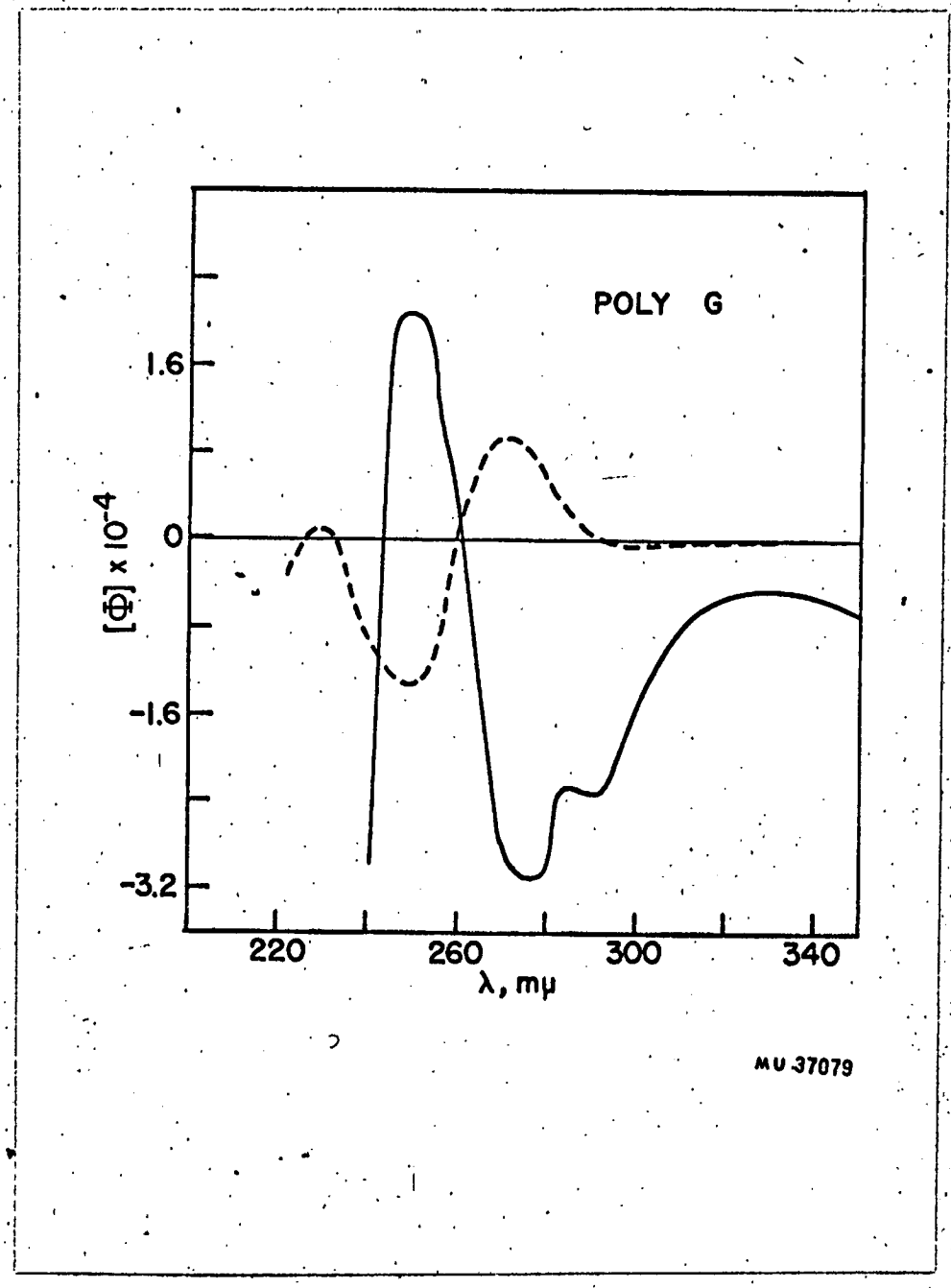


Figure 16. The optical rotatory dispersion of poly G at neutral pH.

——— experimental[196]

– – – – – nearest neighbor calculation

The experimental curve used for poly G had to be obtained in an indirect way. Sarkar and Yang report the ORD of a mixture of poly G and poly C before and after heating to 95°C.[196] Pochon and Michelson have shown that at around neutral pH poly G has an extremely stable secondary structure.[174] Thus the renatured data of Sarkar and Yang almost certainly represent poly C and poly G aggregate. By subtracting the known ORD of poly C we can make an estimate of the ORD of poly G as it exists in 0.1 molar tris, pH 7.5.

The agreement between the experimental and calculated ORD of poly C, poly A and poly U is qualitatively very good. The correct shape of the multiple Cotton effects are reproduced in each case. But in all three polymers, the trough in region of 260 mμ is shifted several mμ to short wavelength from the value predicted from the dimers. In addition, the rotation of poly A and poly C is considerably more intense at wavelengths above 250 than predicted from our calculations. The opposite occurs in poly U, whose rotation is smaller than one would predict from the dimers. In all three curves the longest wavelength crossover is shifted to the blue relative to the calculated value. This shift is 6.2 mμ in poly A, 4.7 mμ in poly C, and 1.8 mμ in poly U. The average of these shifts is 4 mμ and, as we shall see, this is very much smaller than changes which occur when double strands are formed.

The less than perfect agreement between experimental and calculated results for poly A, C, and U can be traced to the breakdown of some of the assumptions used to perform the calculations. The polymer has electrostatic and geometric constraints which are absent in the dinucleoside phosphate. These would tend to make the polymer a more rigid rod than

one would expect from studying dinucleoside phosphates. In addition, it is possible that the structure of homopolymers in solution is very regular. In this case, next-nearest neighbor interactions would always be additive, and this could result in the discrepancies we have observed. Fortunately, for an RNA with a more random base sequence, we might expect far less regularity, and thus next-nearest neighbor interactions should at least partially cancel. It has been pointed out to us that a better approximation than the one we have used would be to use the molar rotation of dinucleoside phosphates as they occur in one polymer to predict the rotation of other RNA's.[91] This would correct for differences in conformation between dimers in a polymer and dimers surrounded by solution. It would also take care of exciton interactions present in the polymer but absent in the dimer. Unfortunately, we do not have sufficient information yet to extract the rotation of more than a few dimers from polymers. But in the future this will be a valuable approach. However, it is worthwhile to keep in mind that the conformation (and electronic environment) of a dinucleoside phosphate contained in a single strand polymer may well depend on the types of residues which surround it. To answer this and related questions, it is most desirable to study polymers or large oligomers of known sequence. Attempts in this direction are just beginning in our laboratory.

The agreement between the experimental ORD curve for poly G shown in Figure 16 and the curve calculated by nearest neighbor methods is nil. We feel fairly certain that this is not due to an inadequacy of our methods. It is well known that poly G is an aggregate which is at least double stranded. Since our calculations are for single strand poly G,

we should be encouraged by the fact that there is no agreement. This means that large optical changes can occur upon going from single to double strand G's. It is interesting to note that the ORD of aggregated poly G shown in Figure 16 is qualitatively similar to the ORD of the pG aggregate.[197] This may reflect similarity in structure between the two aggregates. Both of these aggregate spectra are very different from any other ORD observed for compounds containing only the bases A, U, C, and G. Only in the aggregate spectra is the long wavelength Cotton effect negative.

We are encouraged by the extent of agreement between calculated and experimental ORD of polymers which are known to be single strands. We can be confident that a calculated ORD for an RNA single strand will be close to the experimental values. Thus we can now use ORD as a probe to see if RNA's are single strands under various conditions.

8. Randomness of Sequence in RNA

To calculate the ORD of an RNA, one would like to be able to use equation (12) of Chapter III. But in order to do this one would have to know the number of times each nearest neighbor pair appears in the polymer. While this is not as much information as knowing the sequence of the RNA, we know of no way of obtaining nearest neighbor frequencies accurately for most natural single strand RNA's. For DNA, the precursors dATP, dGPT, dTTP, and dCTP are selective enough so that these may be radioactively labeled, fed one at a time to an organism, and the DNA isolated. Then the DNA is cleaved to give 3' deoxynucleotides. The

location of the radioactivity in these fragments clearly gives all of the nearest neighbor frequencies. But there are no good specific precursors for in vivo RNA synthesis, since the ribonucleoside triphosphates very actively participate in the intermediary metabolism of the cell, and thus the labeling is rapidly randomized. RNA nearest neighbor frequencies have been determined for RNA synthesized in vitro by RNA polymerase.[113] But for most RNA's, the only accurate way we have to learn nearest neighbor frequencies is to determine the base sequence. Thus far this has only been accomplished in the case of the alanine sRNA.[97] For other RNA's, we shall have to be content with approximations.

If we make the assumption that the sequence of RNA is random, we can use equation (13) of Chapter III to calculate the ORD of RNA. By random sequence, we mean only that the equation $X_{ab} = X_a X_b$ (X is mole fraction) is as accurate as our experimental data. We feel that a 5-10% deviation in X_{ab} would have no measurable effect on our ORD calculations. We would like to stress that the assumption of a random sequence in RNA does not contradict the fact that this molecule is carrying information. In fact, a random sequence can, in principle, contain as much information as any other sequence. Thus the assumption of random sequence is not a very great restriction. But is it valid?

There is much experimental data which can be analyzed to give some evidence on the randomness of various RNA's. The quantitative analysis of fragments isolated from a native RNA will permit the frequency of certain sequences to be ascertained. For example, the base composition of mixtures of dimers through tetramers isolated from pancreatic RNAase digestion of R17 RNA is in reasonable accord with what one expects from

a random sequence.[213] But in the monomer peak U/C was 1.38. This ratio in the whole RNA is known to be only 1.04. However, Sinha, Fujimura and Kaesberg found that by their isolation procedure the total amounts of U and C contained in all of their fragments had a ratio of 1.27. This is presumably due to deamination of C to U.[213] We give this example to show that one must be cautious in interpreting some of the data reported for randomness of sequence.

In Table IVa we have collected some of the more recent data on the occurrence of several fragments in pancreatic RNAase digests of RNA's. The values shown in the table represent the amount of fragment isolated experimentally divided by what is expected if the seqeunce is random. It is seen that the monomers and dimers occur in very close to random frequencies. These results break down when the analysis of trinucleosides prepared by pancreatic RNAase digestion is considered. However, it is not known whether to interpret large deviations in the amount trimer as arising from deviations in nearest neighbor frequencies. It could be that these frequencies are still quite random, but there is a strong preference for certain dimers to occur in sequence. This would have no effect on nearest neighbor calculations but, of course, would greatly effect the message coded on the RNA.

It is possible to check the long distance randomness of RNA's by comparing the distribution of chain lengths isolated from a pancreatic RNAase digestion with the distribution calculated from random RNA. Fiers, Lepoutre and Vandendriessche find that there is a large deviation between the amount of monomer found and expected from a digest of yeast RNA.[61] This deviation rapidly goes to zero as the chain length is

TABLE IV

(a) Evidence in Favor of Random Sequences in RNA

Pancreatic RNAase Digestion

RNA	Reference	Amount of Oligomer Found/Amount Expected from Random Sequence					
		C	U	AC	AU	GC	GU
MS2	58	1.09	0.98	1.07	0.87	0.89	0.93
F2	8	1.05	1.01	1.04	1.02	1.00	0.85
Yeast	58	1.30	1.14	1.03	0.86	1.17	0.69
T4 mRNA	8	0.94	1.01	0.86	1.03	1.11	0.87
E.coli rRNA	8	1.03	0.97	0.94	0.72	1.05	0.91
TMV†	186	0.93	0.87	0.86	0.78	0.87	0.93

(b) Evidence in Favor of Non Random Sequences in RNA

Results Adapted from Miura[155]

Sequence*	Nearest Neighbors Found/Calculated from Random Sequence			
	High M.W. yeast RNA	Yeast sRNA	TMV RNA	Rat Liver rRNA
G_g	0.88	1.02	0.96	0.66
G_{u+c}	0.86	0.90	1.10	0.83
G_a	1.37	1.80	1.24	2.07
C_g	0.74	0.84	0.67	0.82
C_{u+c}	1.01	1.10	0.98	0.94
C_a	1.25	1.58	1.32	1.49
U_g	0.74	0.97	0.78	0.99
U_{u+c}	1.02	0.90	1.01	0.82
U_a	1.25	1.88	1.30	1.48

*G_g means G after g, etc.

†We computed the values shown from the base composition of Ref. 256.

increased. Similar experiments with MS2 RNA suggest that the deviation, though smaller, is periodic with chain length. This may represent idiosyncracies in the genetic code. Much better agreement between calculated and experimental chain length distributions have been found by Sinha and his coworkers for R17[213] and M12[212] RNA. In our laboratory, Yolles has found that the deviations from randomness in TMV RNA are also small.[264] Thus there is a large body of results which indicates that the nearest neighbor frequencies of RNA's are close to random.

A challenge to the above results comes from the work of Miura,[155] which is summarized by Egami, Takahashi, and Uchida.[51] The methods used by Miura should, in principle, give most of the nearest neighbor frequencies of an RNA. The procedure is as follows. The RNA is hydrolyzed to completion with T1 RNAase. The amount of G isolated as monomer must have all occured after G in the sequence. Thus, in Miura's notation, we have determined G_g, the G after g. By permitting simultaneous digestion of the RNA by pancreatic and T1 RNAases, we liberate all of the G except that which occurs after A in the sequence. Thus this permits us to obtain G_a. We then use the equation that $G_g + G_c + G_u + G_a$ equals the total amount of G in the polymer. This permits calculation of G_{c+u}. Similar experiments led to other frequencies, such as C_G, C_{U+C}, C_A, etc.

The experimental results of Miura are compared with calculations in Table IVb. The numbers shown represent the ratio of nearest neighbor frequency found divided by the value calculated for a random RNA. It can be seen that in every case, A is found after C, U, or G more than would be expected from a random sequence. In all cases but one, there is less G observed after G, U, or C than would be found in a random se-

quence. If the results shown in Table IVb are correct, it will be very inaccurate to assume that sequences in RNA are random. In addition, it is very hard to reconcile the results shown in Table IVb with those in IVa. We find it hard to believe that the results of Miura represent the true nearest neighbor frequencies. More likely, they arose from a systematic experimental error. Mandeles has found that under normal reaction conditions it is extremely unlikely that T1 RNAase can digest an RNA to completion.[143] Since this enzyme is at least partially specific for single strands, the large double stranded sections which probably occur at random in a large RNA will be invulnerable to enzyme attack. If this occurs, one would always expect the results of U_a, C_a, and G_a to be high. Contrastingly, the results for G_g, C_g, and U_g will be lower than the true values. Since this is exactly the way Miura's results turn out, we suspect that his digestions did not go to completion. Thus, though we shall ignore his results, we feel that with suitable modification his method is still useful for determining nearest neighbor frequencies.

We shall assume in the following sections that the sequences of all RNA's except alanine sRNA are random.

9. ORD of Single Strand RNA

We have calculated the ORD of many single strand RNA's using equation (13) of Chapter III. These computations were performed on an IBM 7094 digital computer using the program shown in Appendix 2. There is no point in presenting most of the data here, since there are very few experiments available for comparison. And, under most conditions, we do not expect RNA to be single stranded. But our calculated data is still of great use for the following reasons. The ORD of an RNA in solution

will arise from interactions among bases on the same strand, and between bases on opposite strands. This means that to analyze an experimental RNA ORD curve, a fairly large number of parameters would be needed. But it can be assumed that the single strand contributions to the ORD do not strongly change when the single strands form a double strand. In this case, our ability to compute the ORD of single strands will be very useful. For we shall be able to subtract the single strand contribution from the ORD in solution. This will permit a simpler analysis of the data which remains. Of course, under some conditions RNA is almost certainly a single strand. If these conditions do not change the single strand stacking, a direct comparison between experiment and calculations will be possible. However, it will not be possible to compare RNA data in urea or ethanol solutions with dimer data in water, since the bases are unstacked in the former two, but stacked in the latter.[38]

Some of the ORD curves we have calculated are shown in Figures 17 through 19. These have been selected to illustrate certain features of single strand RNA ORD. Values of selected peaks and troughs are shown for a wider selection of RNA's in Table V. In Figure 17, we have plotted the ORD expected for three RNA's of random sequence. The first contains only a 1:1 ratio of A to U. Another is a hypothetical RNA with a random sequence and only a 1:1 ratio of G and C. The last is random RNA equimolar in the four normal bases. These three curves show the effect of change in base composition on the ORD of RNA. It is easily seen that if RNA contained A=U and G=C, as occurs in DNA, it would be possible to determine base composition by measuring ORD. One would use the same analysis that has been applied to DNA by Felsenfeld and Hirshman.[57] The

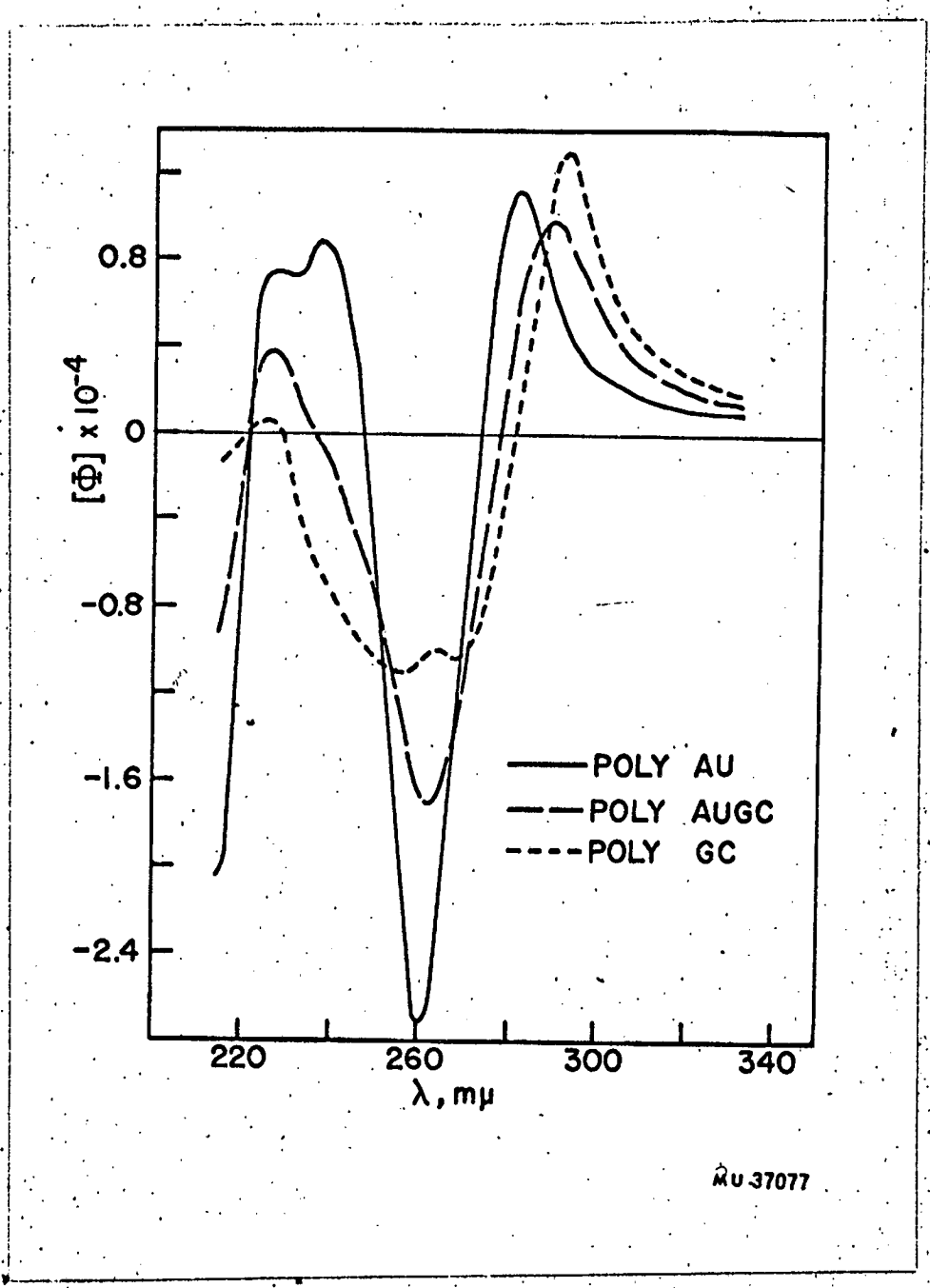


Figure 17. The effect of base composition on the calculated ORD of single strand polynucleotides with random sequence, at neutral pH.

———— poly (AU) 1:1

— — — poly (AUCG) 1:1:1:1

- - - - - -poly (CC) 1:1

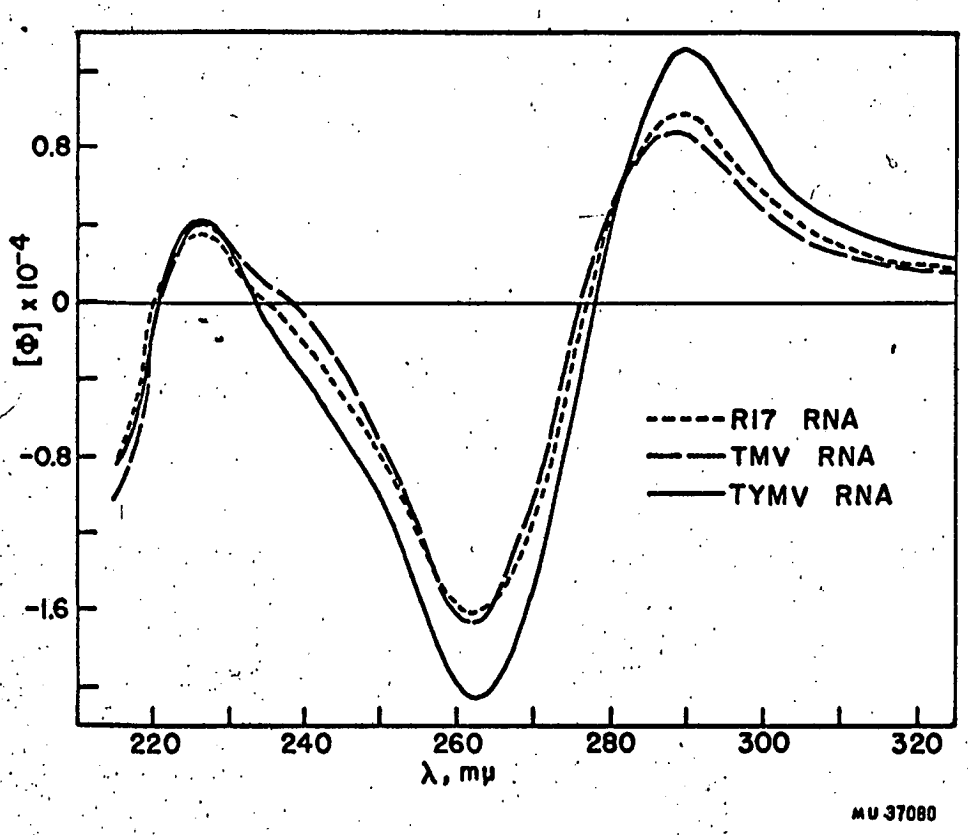


Figure 18. Calculated ORD of RNA's assumed to have a random sequence, and a single strand conformation, at neutral pH.

———— TYMV RNA

— — — TMV RNA

- - - - - -R17 RNA

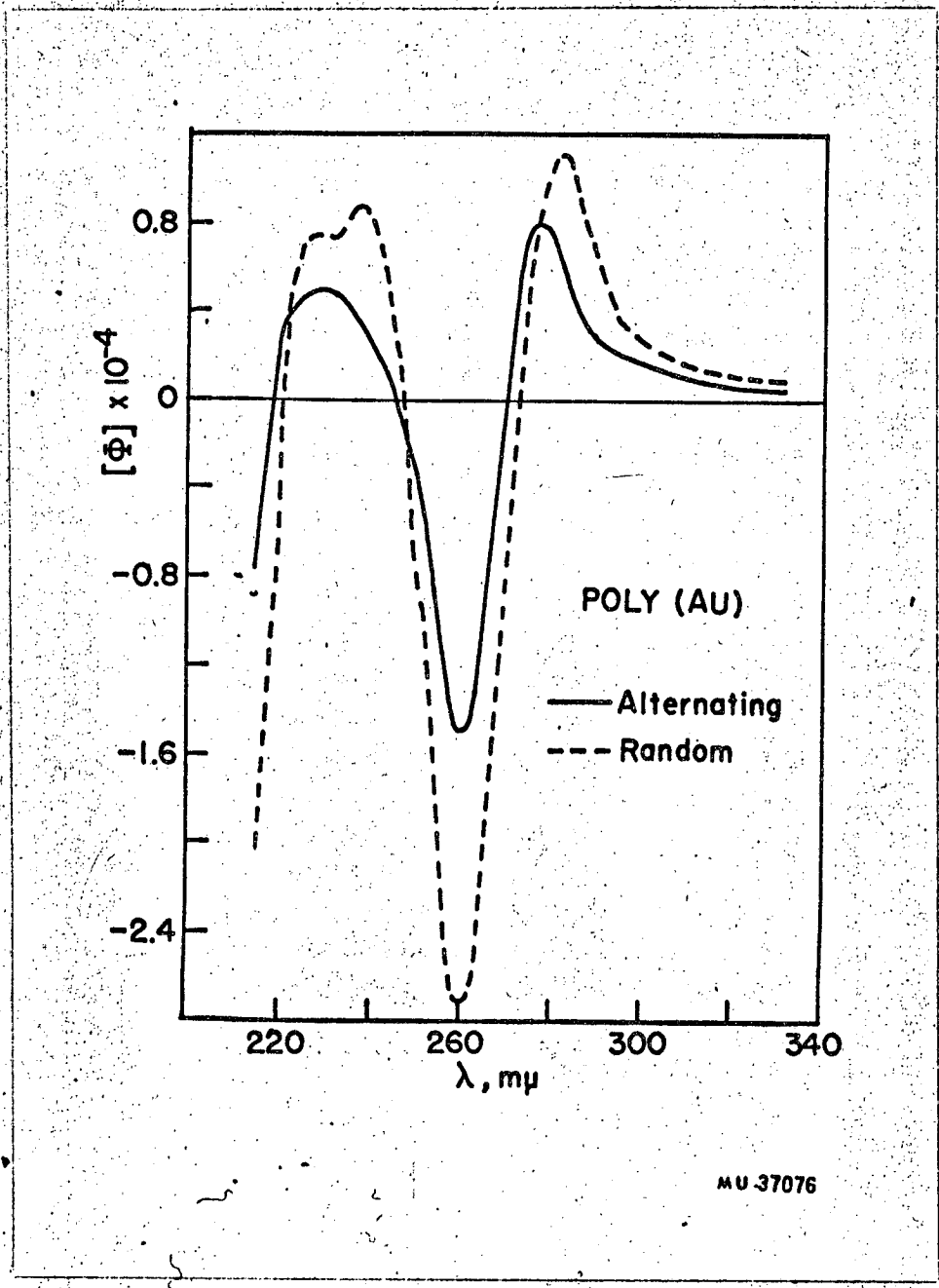


Figure 19. The effect of base sequence on the calculated ORD of single strand polynucleotides, at neutral pH.

———— poly (AU) 1:1, random sequence

– – – – – poly (AU) 1:1, alternating sequence

equality of A and U and G and C does apply in the case of double strand RNA's like reovirus, but the ORD curves we have calculated are for single strands. As we shall show later, the approximation that A=U and G=C is not a bad one for most RNA's, and thus the curves shown in Figure 17 represent fairly good approximations for a typical single strand RNA.

It is of much more interest to the experimentalist to see a curve calculated for a real RNA than the hypothetical base compositions shown above. Thus, in Figure 18, we have plotted the ORD of three RNA's: TYMV, TMV, and R17. It is easily seen that the ORD curves of these three RNA's are reasonably different. The results for TYMV are strongly influenced by the fact that it contains 38% C, more than any other known RNA.[259] For TMV RNA, we have used the base composition given in Haggis et al.[82] To calculate the ORD of R17 RNA, the base composition reported by Simha, Fujimura, and Kaesberg was used.[213] The base composition of R17 RNA is close to equimolar in the four bases, while TMV RNA has slightly more A and U than C and G. From Figure 18 it can be seen that ORD is reasonably sensitive to the changes in base composition among natural RNA's. However, the shape of the curve is fairly similar for most RNA's; the differences among them are mostly quantitative. Since we cannot trust the quantitative details of ORD curves calculated by nearest neighbor methods to better than 10%, even under close to ideal circumstances, it would be unwise at this time to try to determine base composition of an RNA by measuring its ORD. But this approach should prove useful in the future.

ORD is, in principle, able to discriminate differences in sequence. The assumptions we have made above will wash out any of these effects. But for a long RNA, in most cases, differences between sequences will

affect the ORD by miniscule amounts. An exception occurs in the case of RNA's of regularly alternating sequence. These can be compared with random copolymers, and, in the extreme case, block copolymers. Since the single strand contribution to the ORD of alternating copolymers will be of use in understanding the ORD observed by double strand complexes of these molecules, we have calculated the ORD for all of the alternating copolymers which are likely to be studied in the near future. These include poly (AU), poly (GC), poly (AC), poly (AG), poly (CU), and poly (UG). The rotation at selected peaks and troughs for these polymers is given in Table V. In addition, the ORD of poly alternating (AU) is compared with the ORD expected for random poly AU, in Figure 19. From these results one can see that the sequence of a polymer can, in extreme cases, substantially effect its ORD. Larger effects can be expected in shorter polymers where large fluctuations from random sequence might be possible. But measuring the ORD of most large RNA's cannot supply any direct information on their base sequence.

10. Effect of Double Strands on ORD

In this section we would like to review briefly some of the ORD and CD data which have been obtained by other workers for single and double strand homopolymers, and for natural RNA's. The purpose of this review is to convince the reader that the formation of double or triple strands has a marked effect on optical properties. With this background, we will then attempt to analyze the ORD of several RNA's in more detail.

Brahms and Mommaerts have measured the CD of yeast rRNA, yeast sRNA, liver sRNA, TMV RNA, and reovirus RNA.[96] In all of these cases, they

TABLE V

Calculated ORD of Some Single Strand Polynucleotides at pH 7

Substance	Long Wavelength					Short Wavelength		
	λ_p	$[\phi_p]$	λ_o	λ_t	$[\phi_t]$	λ_o	λ_p	$[\phi_p]$
R-17 RNA	289	.97	277	261.5	-1.62	235	226	.36
TMV RNA	288.5	.88	276	261.5	-1.66	238	226.5	.41
T-4 mRNA	287.5	.89	275.5	261.5	-1.85	241	227	.46
Ala-sRNA	291	.97	279	263	-1.10	230	227	.12
Ala-sRNA*	291	.97	278	262.5	-1.14	231	227.5	.18
MS2 RNA	289.5	.96	277	262	-1.60	235	226	.36
Yeast-sRNA	290	.99	277	261.5	-1.55	233.5	226.5	.34
TYMV RNA	290	1.32	278	262.5	-2.04	233.5	226.5	.43
F2 RNA	290	1.01	277	262.5	-1.63	234	226.5	.35
Poly A	282.5	2.07	274	260	-5.80	249	239	2.67
Poly U	285	1.16	276	261.5	-2.08	238.5	228	.44
Poly C	292	3.12	281.5	275	-3.40	-	226	-.08
Poly G	271	.98	259.5	250	-1.30	233	228	.14
Poly(AU)†	278	.79	270	261	-1.49	246	229	.50
Poly(GC)†	293.5	1.05	281.5	260	-1.02	228.5	218.5	.55
Poly(AG)†	288	.28	277	267.5	-.89	246.5	240	.18
Poly(UG)†	291.5	.42	282	272.5	-.60	264	256	.64
Poly(AC)†	287	1.60	277	263	-3.23	237	227.5	1.68
Poly(UC)†	289	1.97	277.5	261.5	-2.43	-	229	-.10
Poly(AUGC)††								
0:0:10:10	292.5	1.31	281	256	-1.10	227	225	.06
1:1:9:9	292	1.24	280.5	260	-1.14	229	226	.13
2:2:8:8	291.5	1.16	279.5	262.5	-1.26	230.5	226	.18
3:3:7:7	291	1.09	278.5	262.5	-1.39	231.5	226	.25
4:4:6:6	290.5	1.03	278	262	-1.54	233	226.5	.32
5:5:5:5	289.5	.96	277	261.5	-1.70	236	226.5	.39
6:6:4:4	288	.94	276	261.5	-1.86	240	226.5	.46
7:7:3:3	286	.93	275.5	261.5	-2.03	242.5	226.5	.53
8:8:2:2	284	.95	274.5	260	-2.23	245	227	.60
9:9:1:1	283	1.03	273.5	261	-2.46	246	227.5	.67
10:10:0:0	283	1.12	272.5	261	-2.71	247	237.5	.88

*Calculated by assuming the base sequence is random.
†Alternating sequence.
††Random sequence.

observe a single dichroic band with a maximum at 265 mμ. This would correspond to an ORD consisting of a positive Cotton effect with a cross-over at 265 mμ. All of the RNA's measured had about the same rotational strength. Since reovirus is known to be double stranded, this implies that the other RNA's are probably mostly double strands under the conditions in which these experiments were performed. Other workers have measured the ORD of mixed sRNA's under equivalent conditions, and their findings are consistent with the above results.[116,55,105] In addition, Lamborg and Zamecnik have found a much stronger Cotton effect centered around 198 mμ in E. coli sRNA.[115] The ORD found by Samejima and Yang for yeast RNA is also consistent with these results, although they did not penetrate quite as far into the UV.[192] It is of interest to note from the work of Samejima and Yang that the ORD of RNA (presumably mostly double stranded) is very different from the ORD of native DNA.[192] Whether this is due to different double strand conformations or to the effect of the extra assymmetry of the ribose sugar is not known. But recent work of Jaskunas has shown that the ORD of poly r(AU) is quite different from the results observed by Samejima and Yang for poly d(AT).[105]

Brahms has studied the CD of poly A, poly U and various complexes.[18] One of his more interesting results is that the CD of poly AU random copolymer is a little smaller than poly (A+U). This may be due to more double strands in the latter than the former. Some of these results are summarized below. Except for double strand poly A, the conditions are pH 7.4, 0.1 M NaCl, at temperatures between 2 and 14°C. The rotational strength, R_{ba}, is defined by the following equation,[237] where in this case the integration is carried out over the longest wavelength CD band.

ϵ_L and ϵ_R are the extinction coefficients of left and right hand circularly polarized light.

$$R_{ba} = 2.295 \times 10^{-39} \int [(\epsilon_L - \epsilon_R)/\nu] d\nu$$

Circular Dichroism of Polynucleotides[18]

Compound	$R_{ba} \times 10^{+40}$ cgs
poly (A+U)	23
poly AU (1:1)	20
poly AU (2:1)	17
poly A + poly U separate, 3°C	32
poly (A+2U)	19
poly A + 2 poly U separate, 3°C	28
poly A alone	33
double strand poly A	64

From these data, one can conclude that as a general rule multistrand helices with different bases have relatively low rotational strengths. The opposite effect is observed upon forming double strand poly A. In any case, the CD shows substantial changes upon changing from single to double strand polynucleotides.

Similar effects have been observed by Sarkar and Yang in their study of the ORD of poly A, poly U, and mixtures of the two.[195] They find that upon mixing poly A and poly U in equal amounts there is a marked shift to the blue of the long wavelength Cotton effect. A slight further shift and broadening is observed when a second strand of U is added to make the triple strand poly (A+2U). Similar studies have been carried out on

poly C and poly G.[196] Sarkar and Yang started with a 1:1 complex of poly G and poly C synthesized by Haselkorn.[90] This presumably contains short strings of G bound to a longer strand of poly C. Upon heating to 95°C, a very large change occurs in the ORD which is maintained after re-cooling to room temperature. We have discussed this previously, and attributed the effect to the formation of single strand poly C and aggregated poly G. Sarkar and Yang have found that the poly (I+C) complex shows a strong peak at 254 mμ not present in either single strand. Similar effects have been observed in poly (A+2I) and poly (3I) relative to their single strands.[196] This all suggests that ORD is very sensitive to the formation of multiple strand helices.

At low pH, both poly A and poly C form double strand helices. Holcomb and Tinoco have shown that at pH 4.5 the ORD of poly A is very different from that at pH 7.[95] The long wavelength trough shifts to shorter wavelength. Similar studies by Fasman, Lindblow and Grossman have shown that there is a very large change in the ORD of poly C when the pH is lowered to 5.[54] In this case the ORD change is accompanied by a substantial change in UV spectrum, since the absorption curve of protonated C is quite different from the uncharged base. Unlike poly A, the trough of double strand poly C is shifted to the red when compared with the single strand.

11. Dependence of the ORD of RNA on Salt Concentration

To obtain experimental data to compare with our calculated ORD for single strands, we had to find conditions under which RNA was likely to exist as a single strand. But these conditions must not be too strongly

denaturing lest they unstack the bases of the single strand formed. One possibility was to study RNA's at elevated temperatures in the hope that a temperature could be found which would break hydrogen bonds but leave single strands still relatively stacked. Such experiments are now in progress in our laboratory.[105] In order to interpret them, one will have to know the temperature dependence of the ORD of all 16 dinucleoside phosphates and the four monomers. These data were not available to us at the time this work commenced.[37]

Another possibility is to measure the ORD of RNA in the limit of low salt concentrations. This will presumably favor single strands over double strands because of the coulombic repulsion between the phosphates on opposite strands. Shildkraut and Lifson have studied the salt dependence of the temperature transition of DNA.[199] They find that at low ionic strength (0.01 to 0.20 molar KCl or NaCl) T_m is given approximately by the equation below.

$$T_m = 16.6 \log (M) + 0.41 (G\text{-}C) + 81.5$$

(M) is the molar concentration of salt. (G-C) is the mole fraction of G-C pairs. This takes into account the dependence of T_m on base composition. A model which takes into account only interaction of charges on phosphates on opposite strands gives fairly good agreement with the above empirical equation. Thus Shildkraut and Lifson indicate that ionic strength effects on single strands are much smaller than double strands. This seems reasonable since in most single strand conformations the distance between neighboring phosphates remains the same.[199] Thus the coulombic interaction can effect only next-nearest neighbor and more distant interactions.

Before we can interpret ORD data taken at low salt concentrations, we must be sure that there are no major salt effects on the ORD of single strands. Jaskunas has shown that the ORD of ApA and CpC in the absence of salt (salt was removed by a Bio-Rad P-2 column) at most wavelengths is identical with the ORD measured by Warshaw for dimers at 0.1 ionic strength.[252,105] This is in agreement with our results on the ORD of trinucleoside diphosphates. The molar rotation of several trimers at selected peaks and troughs is summarized in Table VI. The data in salt were obtained as described in Chapter III. The salt free data were from samples which had been dialyzed continuously against water for two days, lyophilized, redissolved in water, redialyzed, lyophilized, and then taken up in water at about 1/5 the volume they started at. The ORD data for the salt free solutions is not quite so accurate as our normal ORD, since these curves were run on a less sensitive scale. But we think it quite fair to conclude from the data in Table VI that there is no salt dependence of the ORD of trinucleoside diphosphates. Tomlinson has studied the salt dependence of the ORD of poly A at pH 7.[240] He finds that there is a 5-1/2% increase in the magnitude of the long wavelength trough upon going from 0.15 molar salt to salt free solution. But there is no change in the shape of the ORD curve. This is a relatively small effect, and we shall assume it represents an upper limit to the type of salt effects expected with single stranded RNA.

The properties of TMV RNA are known to be strongly dependent on the salt concentration. Boedtker has shown that at ionic strengths of below 0.06 this RNA behaves as a tightly coiled highly flexible polymer chain.[16] The viscosity and radius of gyration from light scattering measurements

TABLE VI

Salt Dependence of Optical Rotatory Dispersion of Trinucleoside Diphosphates

Compound	Wavelength mμ	Molar Rotation x 10^{-4} 0.1 ionic strength	Molar Rotation x 10^{-4} No salt added
ApApU	280	1.10	1.09
ApApU	260	-2.92	-2.95
GpGpC	290	.45	.55
GpGpC	250	-.76	-.72
GpApU	280	.79	.88
GpApU	260	-1.14	-1.20
ApGpU	290	.17	.21
ApGpU	270	-.65	-.67
ApGpU	250	.13	.30
GpGpU	275	.39	.27
GpGpU	250	-.55	-.44

show strong dependence on ionic strength. But above 0.06 ionic strength, the properties of TMV RNA seem to become relatively independent of ionic strength. In a later paper, Boedtker concluded that the hydrodynamic behavior of RNA suggests that it is essentially a random coil.[17] The T_m is very dependent on ionic strength, and sedimentation changes coincide with hypochromicity. This suggests that when secondary structure is formed there are large changes in the spatial arrangement of the molecule. Boedtker emphasizes that with the range of conditions studied RNA is neither completely helical nor completely disordered.[17] Most of her results seem to be consistent with the idea that TMV RNA consists of relatively ordered regions spaced by random ones.

Boedtker found that Mg^{++} was extremely effective in stabilizing the secondary structure of TMV RNA. She estimated that it is 25,000 times more potent than NaCl. But Mg^{++} is not necessary for the formation of secondary structure.[17] If RNA is a loosely stacked helix at very low ionic strength, and a compact helix with double strand hairpin loops at high ionic strength, most of the data of Boedtker can be explained. A flexible stacked helix would hydrodynamically behave like a stiff coil. And, if single strands must fold back on one another to form a double strand, the optical properties will be closely linked to the hydrodynamic observables.

It has previously been shown that metal ions can have pronounced effects on the long wavelength optical rotation of TMV RNA.[89] The results of Boedtker just described encouraged us to study the ORD of TMV RNA as a function of salt concentration. All of the experiments with TMV RNA described in this section were performed by Jaskunas.[105] Our

contribution is to the analysis of the results. We shall outline the experimental methods used when they are substantially different from the procedures already described. TMV RNA was a present from Dr. Mandeles. We are very grateful for samples from some of his highest purity preparations.[143] Salt was removed by a complex dialysis procedure which was designed to make sure that Mg^{++} was removed. Thus frequent dialysis was carried out against various concentrations of EDTA before most of the EDTA was removed by dialysis against dilute solutions. The final result, which we shall call salt free TMV RNA, had been equilibrated against 4×10^{-5} molar EDTA. This had a pH of 6.3. This represents a compromise between lower salt concentrations and the fact that it would have been difficult to lower the salt concentration further without also decreasing the pH. Solutions at higher salt concentrations were made by adding salt to the salt free solution. The dialysis procedure apparently had no deleterious effects on the RNA, since measurements made before and after in the same salt concentration were quite similar.[105] Concentrations were determined by using the extinction coefficient per residue of $1.00 \times 10^{+4}$ in salt free solution,[89] and $0.73 \times 10^{+4}$ in the salt solution.[16] The ORD measurements described here were made at room temperature. Recently these results were extended to a broad range of conditions.[105] Studies by Jaskunas suggest that the conditions used here represent close to fully structured RNA (in salt), and close to the minimum amount of double strand obtainable (no salt).[105]

The ORD of TMV RNA in no salt (4×10^{-5} M EDTA) at pH 6.3 is compared with the results obtained for this RNA at pH 6.5 in the presence of 0.15 M KCl ($+ 10^{-4}$ M EDTA) in Figure 20. It is obvious that salt has a marked

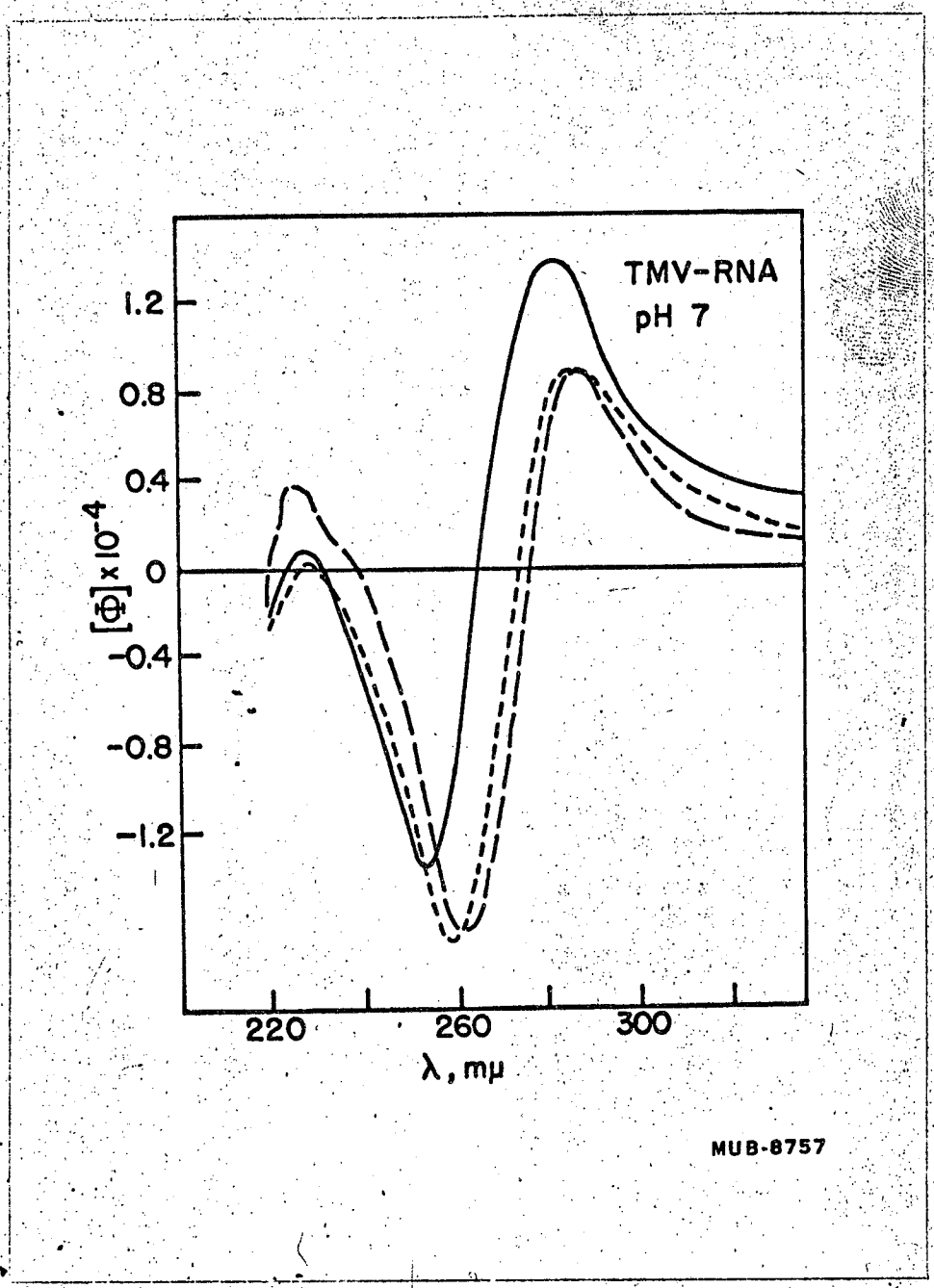


Figure 20. Effect of salt concentration on the optical rotatory dispersion of TMV RNA, at neutral pH.

———— experimental results in the presence of .15 molar salt

— — — calculated by nearest neighbor methods

- - - - - -experimental results in the absence of salt

effect on the ORD of TMV RNA. The crossover in the presence of salt is shifted many mμ to the blue. In addition, similar shifts are noted for the long wavelength peak and trough. Thus we can conclude that the conformation of TMV RNA is very dependent on the salt concentration.

The ORD of TMV RNA was calculated from the dimers and monomers by the methods described earlier in this chapter. This result is shown in Figure 20 along with the two experimental curves. The agreement between calculation and the salt free data is excellent at most wavelengths. Except for very short wavelengths, the calculated curve is much closer to the no salt data than the ORD of TMV RNA in salt. Thus we feel that TMV RNA is probably mostly a single strand stacked helix in the absence of salt. The quantitative agreement between TMV RNA and the calculated curve is difficult to assess. But we are very pleased that the shape of the ORD curve in the absence of salt is reproduced so well by nearest neighbor methods.

Similar results have been obtained for several other RNA's.[105] The agreement between calculations and experiments are best for TMV RNA, but the other cases are comparable. It is certainly fair to expect that some, if not most, RNA's will still maintain some double helical segments even at the lowest salt concentrations we can reach. We are afraid to lower the pH any more lest A-A and C-C hydrogen bonded pairs start to form. This will have an unpredictable effect on the ORD, since each tends to shift the curves in the opposite direction.

12. ORD Difference Curves

We suspected that the change in the ORD of TMV RNA upon adding salt was due to the formation of A-U and G-C pairs. These would arise from folding of the polynucleotide chain into complementary double stranded sections. Thus we would like to be able to estimate the ORD change expected for the formation of a double strand. In principle, this can be done by using nearest neighbor calculations analogous to those described for single strands in Chapter III. For double strands, the interaction we must consider consists of two effects. The first is any interaction between the two hydrogen bonded bases. Since they are in the same plane, this interaction would be small. Most of the contribution to the ORD of a double strand will come from the interaction of vertically stacked base pairs. We know how to account for the interaction of stacked bases in the same strand, but we have very little information of the effect of stacking of one base with adjacent bases on the other strand. If we lump all of these contributions together, there are ten basic interactions which must be determined before the ORD of a double strand can be calculated. They are

1. $\begin{matrix}\text{A-U}\\ \uparrow\downarrow \\ \text{A-U}\end{matrix}$ 2. $\begin{matrix}\text{A-U}\\ \uparrow\downarrow \\ \text{U-A}\end{matrix}$ 3. $\begin{matrix}\text{U-A}\\ \uparrow\downarrow \\ \text{A-U}\end{matrix}$

4. $\begin{matrix}\text{G-C}\\ \uparrow\downarrow \\ \text{G-C}\end{matrix}$ 5. $\begin{matrix}\text{G-C}\\ \uparrow\downarrow \\ \text{C-G}\end{matrix}$ 6. $\begin{matrix}\text{C-G}\\ \uparrow\downarrow \\ \text{G-C}\end{matrix}$

7. $\begin{matrix}\text{A-U}\\ \uparrow\downarrow \\ \text{G-C}\end{matrix}$ 8. $\begin{matrix}\text{G-C}\\ \uparrow\downarrow \\ \text{A-U}\end{matrix}$ 9. $\begin{matrix}\text{A-U}\\ \uparrow\downarrow \\ \text{C-G}\end{matrix}$ 10. $\begin{matrix}\text{C-G}\\ \uparrow\downarrow \\ \text{A-U}\end{matrix}$

where arrows indicate the direction of the polynucleotide chains, and lines represent hydrogen bonded interactions. Each of these interactions could be determined in principle by measuring the ORD of a suitably chosen oligomer complex or double strand RNA. Matters can be simplified somewhat if we subtract the contribution from single strands. Then we can write that the ORD of a double strand is equal to the ORD of its single strands plus contributions from the interactions. One would need to know the frequency of each double strand interaction in order to specify the ORD.

At the present time the approach outlined above is beyond our means. There are simply not enough ORD data available on compounds of well defined sequence. Thus we cannot yet compute most of the changes which should occur when double strands are formed. But Sarkar and Yang have studied the ORD of poly (A+U), poly A and poly U.[195] Thus, from their results, we can estimate the ORD change upon forming an A-U base pair. (This corresponds to interaction 1 in the above list.)

$$\Delta[\phi]_{AU}(\lambda) = [\phi]_{\text{poly (A+U)}} - 1/2\,([\phi]_{\text{poly A}} + [\phi]_{\text{poly U}}) \quad (11)$$

Sarkar and Yang have also studied the ORD of double strand poly (G+C).[252] We have mentioned ORD data for single strand poly C previously. And while no experimental curves exist for single strand poly G, we have calculated its molar rotation as described previously. Thus we can compute the ORD change for forming a G-C pair according to interaction number 4.

$$\Delta[\phi]_{GC}(\lambda) = [\phi]_{\text{poly (G+C)}} - 1/2\,([\phi]_{\text{poly C}} + [\phi]_{\text{poly G}}) \quad (12)$$

$\Delta[\phi]_{AU}$ and $\Delta[\phi]_{GC}$ are plotted in Figure 21.

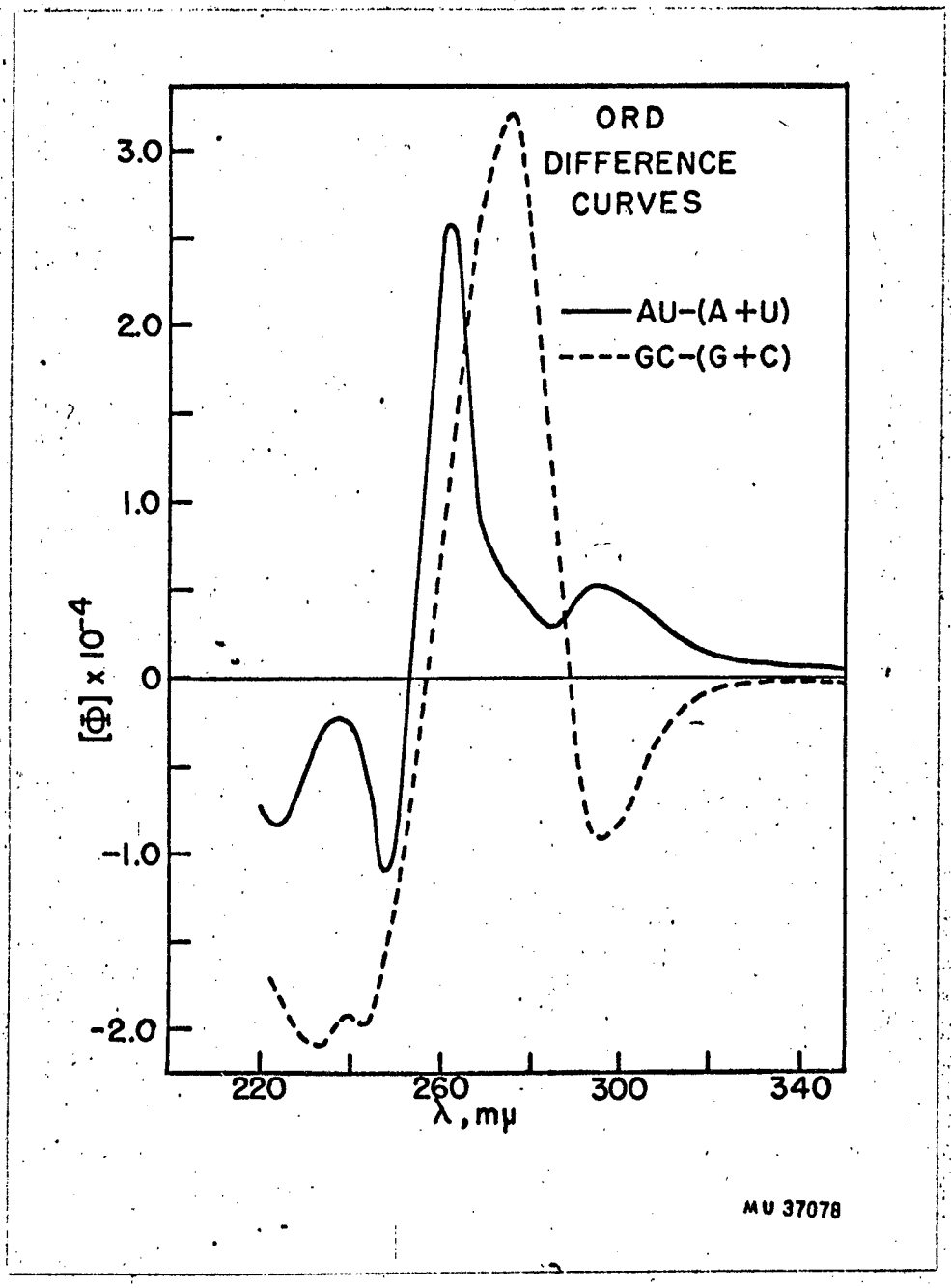


Figure 21. ORD difference curves for the formation of base paired double stranded RNA from single strand RNA. Details are explained in the text.

——— poly $(A+U)^{195}$ - 1/2 (poly A + poly U)

– – – – – poly $(G+C)^{196}$ - 1/2 (poly C + poly G)

To calculate the ORD of double strand TMV RNA using equations (11) and (12), we have to assume that equation (11) is representative of the average ORD change upon forming an A-U pair, and equation (12) is the average ORD change for forming a G-C pair. These are relatively gross assumptions, but it is the very best we can do at present. Very recent data have shown that the ORD change for interaction 1 is the same as for the average of 2 and 3.[105] This lends credence to our above assumption, but we would still like to obtain more experimental data.

Even with the above assumption, we cannot calculate the ORD of TMV RNA unless we know the number of A-U and G-C pairs. There are more A's and U's in TMV RNA than G's and C's.[256] This would tend to favor the formation of more A-U pairs, but results we have presented earlier suggest that the G-C interaction is stronger. Thus, it is a reasonable assumption that TMV RNA has roughly as many A-U as G-C pairs. But this still does not tell us how many there are. We think it would be foolish to guess at this figure. So, instead, we will compare the change in ORD upon forming an average base pair with the changes that occur in the ORD of TMV RNA when salt is added. The ORD change for an average base pair will be just the average of $\Delta[\Phi]_{AU}$ and $\Delta[\Phi]_{GC}$. This is plotted in Figure 22 along with the difference between the ORD of TMV RNA in salt and the calculated single strand ORD. In addition, the change in the observed ORD of TMV RNA when salt is added is shown at the bottom of Figure 22. The similarity in shape among all three curves is evidence that, in the presence of 0.15 molar salt, TMV RNA is partially double stranded. The differences in magnitude between the two bottom curves in Figure 22 suggests that even under our low salt conditions TMV RNA has still formed some double strand segments.

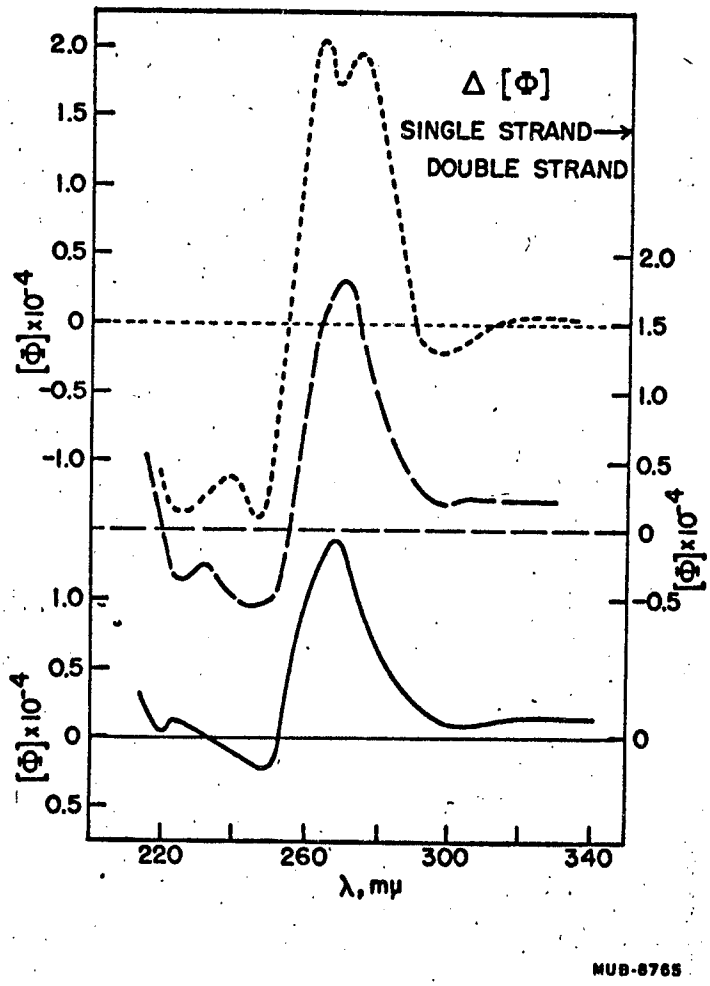


Figure 22. ORD difference curves for the formation of base paired double stranded RNA from single strand RNA. Details are explained in the text.

———— measured ORD of TMV RNA in salt - measured ORD in the absence of salt

— — — measured ORD of TMV RNA in salt - calculated ORD of single strand TMV RNA

- - - - - -average ORD change upon forming a completely double strand RNA with an equal number of A-U and G-C pairs. This is the average of the two curves shown in Figure 21.

It will be very interesting, once the structure of RNA in solution is better understood, to see whether the rough approximations we have used in this section have any validity. For example, we have ignored the effect of G-G interactions, even though they have been demonstrated to be of great importance in small oligomers. And we have assumed that double strands are formed in RNA, but no triple strands were considered. We have restricted ourselves to A-U and G-C pairs at pH's which are dangerously close to the pK's of poly A and poly C in low salt, although a study of ORD versus pH indicates that we are probably still safely close to pH 7.[105] Lastly, we have ignored the fact that if double strands are formed by hairpin loops there must be a loss in stacking for bases in the hairpin. We list all of these omissions to avoid giving the impression that we have clearly demonstrated that TMV RNA contains segments forming double strands in high salt, but not at low salt. Our data is consistent with such an interpretation, but by no means proves it.

13. Base Composition of RNA

The success we have had in predicting the ORD of TMV RNA encouraged us to calculate the ORD of many single strand polynucleotides. Some of the attempts to do this have already been discussed. We wanted to be able to prepare a table of values which would approximate the ORD of any given RNA. But, in order to do this, we were faced with a serious problem. Unlike DNA, where A=U and G=C, RNA is usually thought to have no such regularities in base composition. Thus to define the base composition of RNA, there are four variables and only one constraint. This means that it is impossible to prepare a convenient table of values without

calculating an uncomfortably large number of curves. Thus we decided to collect a large body of base composition data to see if there were any accidental correlations which would simplify our task.

The base composition for 112 RNA's is collected in Table VII. These were determined by many different workers using a variety of methods. The precision of the data varies from better than a few percent,[12] to values in which only one or two determinations were made. The accuracy, as always, is suspect since very few results agree from laboratory to laboratory. We have tried to use the most recent data available for most of these RNA's. The purity of the RNA's in Table VII varies from over 90% to unknown and undetermined. We have tried to be as objective as possible in collecting this set of data from the literature. This list is by no means a complete survey of the available data. Instead, it represents what we hope was a random sampling. We include 27 sRNA's, 21 rRNA's, 33 mRNA's or viral RNA's, 3 nuclear RNA's, and 28 whole cell RNA's. This distribution was dictated by chance, rather than design, but it is fairly random .

To see if there were any systematic relationships governing the base composition of RNA, we plotted the mole fraction of one base against the mole fraction of another base. Four of these plots are shown in Figures 23 through 26. In addition, we tried plotting mole fraction of two bases against the sum of the mole fractions of two other bases, using triangle paper. These plots are illustrative, but the same conclusions can be drawn from the data reproduced in this chapter. If the base composition of RNA was strictly random we would expect a great deal of scatter in the plots shown in Figures 23 through 26. This is observed.

TABLE VII

Base Composition of Various RNA's

RNA	A*	U†	C	G††	Reference
1. Yeast alanine sRNA	.106	.200	.306	.386	97
2. Mixed yeast sRNA	.211	.220	.264	.307	104
3. TMV	.289	.277	.191	.243	256
4. R17	.231	.253	.249	.263	213
5. MS2	.228	.252	.249	.271	226
6. T4 mRNA	.313	.302	.177	.208	8
7. Wheat Germ sRNA	.228	.218	.236	.316	75
8. Wheat Germ high M.W.RNA	.266	.282	.182	.269	75
9. TYMV	.230	.220	.380	.170	259
10. HeLa cell nuclear RNA	.196	.137	.353	.314	84
11. Calf thymus primed	.279	.291	.218	.213	102
12. M.lysecdeikticus primed	.150	.150	.343	.357	102
13. Yeast serine sRNA	.208	.225	.277	.287	191
14. F2	.222	.251	.268	.259	12
15. FR	.249	.237	.243	.271	12
16. 70S ribosomes	.255	.218	.219	.308	12
17. Chinese cabbage 4S	.169	.199	.328	.308	148
18. Chinese cabbage 16S	.190	.263	.220	.325	148
19. Chinese cabbage 24S	.201	.207	.248	.347	148
20. Cucumber virus	.260	.300	.190	.260	263
21. Polio virus	.300	.270	.190	.260	263
22. SBMV	.260	.250	.230	.260	263
23. Z1K/1 virus	.236	.283	.242	.239	12
24. ZJ/1 virus	.243	.282	.237	.238	12
25. Z6 virus	.251	.237	.240	.272	12
26. ZS/3 virus	.248	.234	.238	.280	12
27. Z1/3 virus	.248	.234	.249	.269	12
28. α 15 virus	.247	.240	.250	.263	12
29. Tobacco necrosis virus	.280	.249	.221	.250	172
30. Tomato virus	.275	.245	.204	.275	172
31. 7S virus	.238	.258	.256	.246	176
32. EMC virus	.273	.253	.235	.235	176
33. Reovirus-3	.297	.305	.193	.205	176
34. Calf Liver	.195	.164	.291	.350	172
35. Calf Pancreas	.141	.134	.237	.487	172
36. Rat Liver	.191	.208	.266	.335	172
37. Carp Muscle	.164	.180	.311	.344	172
38. Sea Urchin eggs	.223	.207	.274	.296	172
39. Clostridium perfringes	.260	.193	.245	.302	172
40. Bakers yeast	.264	.237	.198	.301	172
41. Brewers yeast	.255	.245	.235	.265	172
42. Human liver	.115	.126	.316	.441	172

TABLE VII (Cont.)

RNA	A*	U†	C	G††	Reference
43. Ox liver	.171	.217	.339	.273	52
44. Ox kidney	.197	.202	.334	.267	52
45. Rat kidney	.194	.204	.307	.295	52
46. Sea Urchin embryo	.226	.208	.272	.294	52
47. Yeast	.254	.274	.226	.246	52
48. S. marescens	.203	.241	.243	.312	52
49. E. coli	.253	.212	.247	.288	52
50. M. phlei.	.209	.213	.271	.308	52
51. T2 specific mRNA	.280	.350	.160	.210	82
52. T2 specific sRNA	.270	.350	.160	.220	82
53. Rat liver nuclear RNA	.180	.270	.330	.220	82
54. Rat liver sRNA	.200	.220	.290	.290	82
55. Rat liver rRNA	.190	.200	.290	.320	82
56. E. coli rRNA	.250	.210	.230	.310	82
57. E. coli sRNA	.200	.190	.290	.310	82
58. Rabbit liver	.193	.199	.282	.326	106
59. Sheep liver	.204	.144	.258	.394	106
60. Sarcina Lutea	.167	.220	.329	.284	106
61. Potato X virus	.344	.214	.228	.214	106
62. Turnip crinkle virus	.276	.223	.246	.256	146
63. Gramicidin S mRNA	.224	.192	.301	.283	201
64. B.megaterium membrane	.244	.228	.214	.314	266
65. B.megaterium sRNA	.185	.241	.262	.312	266
66. Rous Sarcoma 64S RNA	.251	.224	.242	.283	183
67. Rous Sarcoma 4S RNA	.213	.193	.266	.328	183
68. Chick 28S RNA	.172	.161	.307	.360	183
69. Chick 18S RNA	.218	.207	.264	.311	183
70. Chick 4S RNA	.187	.200	.282	.331	183
71. B.stearothermophilus rRNA	.259	.175	.220	.345	145
72. B. stear. purified RNA	.205	.211	.261	.315	145
73. Rat liver 28S RNA	.178	.170	.320	.330	223
74. Rat liver 18S RNA	.198	.180	.298	.324	223
75. E. coli 23S RNA	.242	.213	.223	.321	223
76. E. coli 18S RNA	.255	.210	.210	.325	223
77. Rabbit reticulocyte rRNA	.196	.156	.295	.353	23
78. Rabbit reticulocyte mRNA	.280	.182	.275	.240	23
79. Pseudomonas	.219	.217	.281	.283	165
80. Newcastle disease virus	.238	.294	.230	.238	47
81. Histone bound RNA	.316	.427	.104	.153	101
82. Tyrosine sRNA-Brown	.212	.231	.274	.283	22
83. E. coli high pH heavy RNA	.238	.204	.274	.283	81
84. E. coli high pH 16S RNA	.249	.209	.270	.272	81
85. E. coli high pH 10S RNA	.289	.220	.251	.246	81
86. BMV small RNA	.255	.246	.203	.296	15

TABLE VII (Cont.)

RNA	A*	U†	C	G††	Reference
87. EMV medium RNA	.252	.263	.205	.280	15
88. EMV large RNA	.282	.254	.194	.270	15
89. Rat cortex low salt	.198	.198	.283	.322	142
90. Rat cortex high salt	.190	.179	.304	.328	142
91. Rat cortex high temp.	.174	.173	.302	.350	142
92. Rat cortex high temp.rRNA	.181	.175	.311	.333	142
93. Rat cortex high salt rRNA	.186	.187	.314	.313	142
94. Cortical neurons	.192	.190	.341	.276	142
95. Deiters nuclei	.227	.194	.314	.265	142
96. Glia-deiters	.268	.189	.254	.270	142
97. Cortex sRNA	.198	.197	.283	.322	142
98. Brain microsomal RNA	.181	.197	.297	.325	142
99. Guinea Pig liver sRNA	.162	.222	.270	.297	22
100. Trout sRNA	.195	.213	.300	.292	22
101. Euglena sRNA	.178	.196	.292	.312	22
102. Pea seedling sRNA	.193	.209	.301	.297	22
103. T.pyriformis sRNA	.186	.210	.289	.302	22
104. Blowfly larva sRNA	.178	.214	.302	.294	22
105. Valine sRNA-Tada	.206	.216	.278	.280	22
106. Alanine sRNA-Tada	.197	.231	.269	.289	22
107. Glycine sRNA-Zachau	.198	.239	.258	.292	22
108. Leucine sRNA-Zachau	.205	.237	.249	.294	22
109. Tyr.+Ser. sRNA-Zachau	.209	.255	.236	.284	22
110. Valine sRNA-Holley	.191	.239	.275	.295	22
111. Tyrosine sRNA-Holley	.217	.215	.267	.302	22
112. Valine sRNA-Zamecnik	.196	.222	.251	.331	22

*Including methyl A
†Including T, dihydro U, ψ
††Including I, methyl I, methyl G, dimethyl G

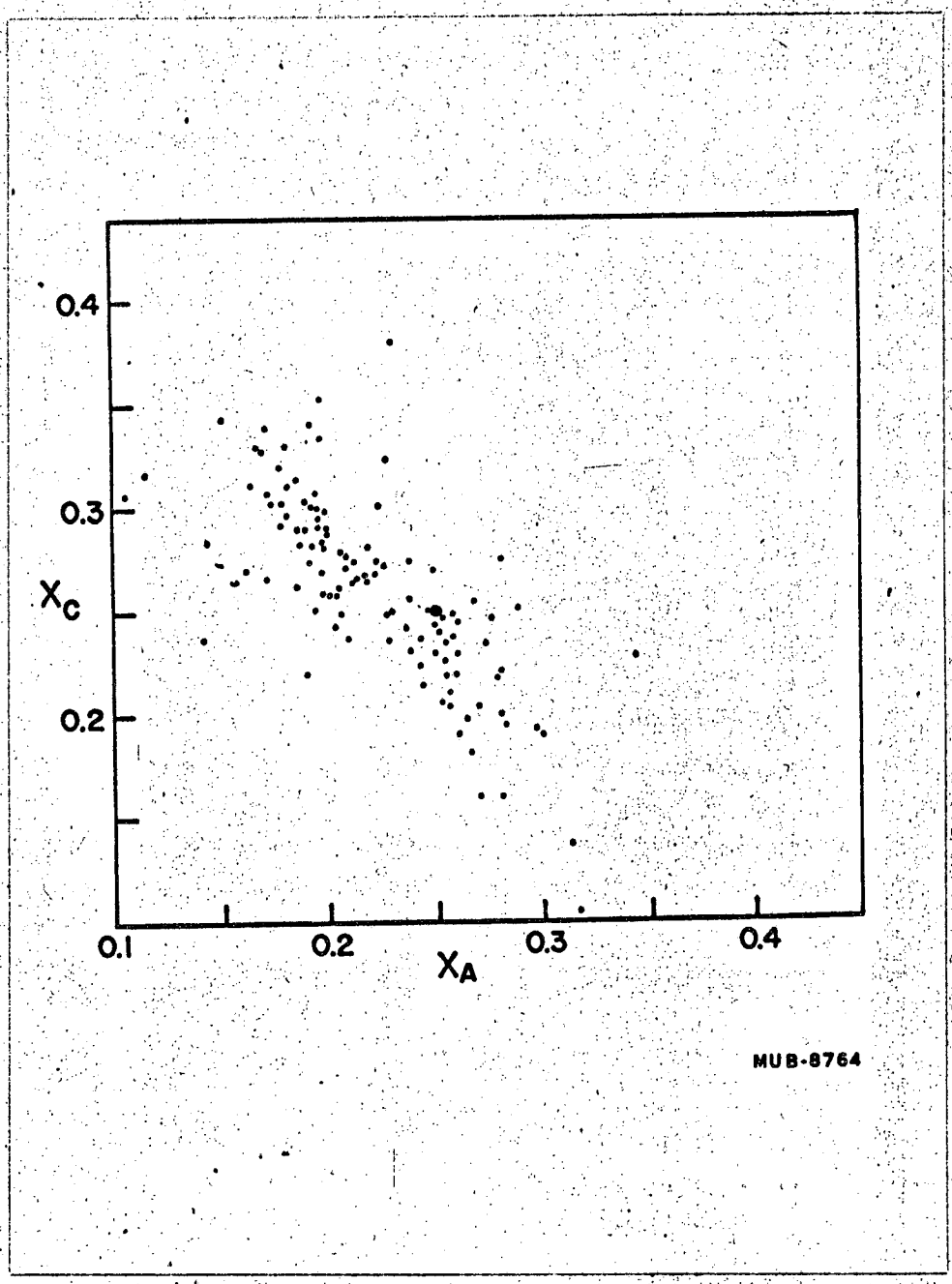


Figure 23. The base composition of various RNA's. A plot of the mole fraction of adenosine against the mole fraction of cytidine for the 112 RNA's listed in Table VII.

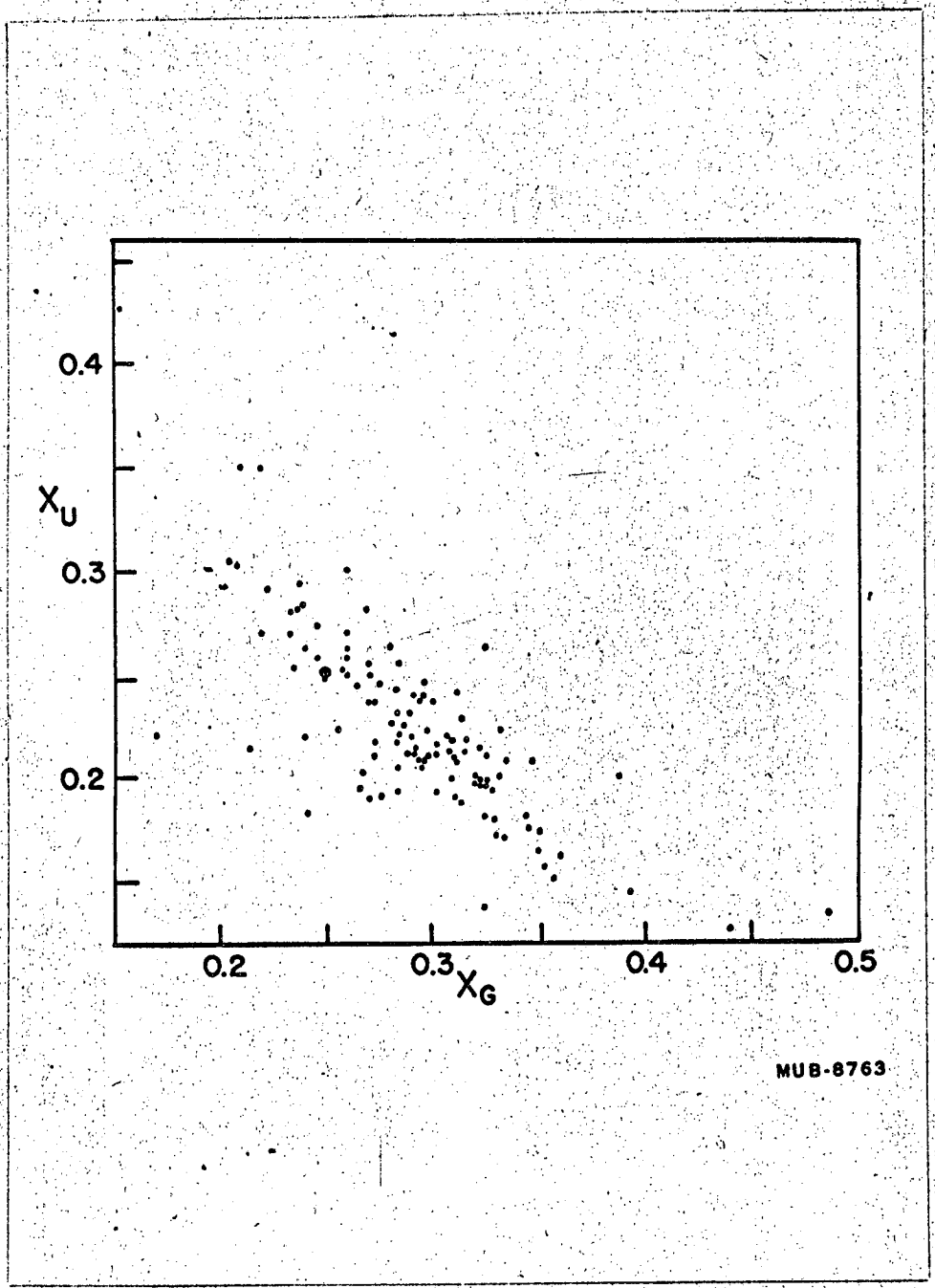


Figure 24. The base composition of various RNA's. A plot of the mole fraction of guanosine against the mole fraction of uridine for the 112 RNA's listed in Table VII.

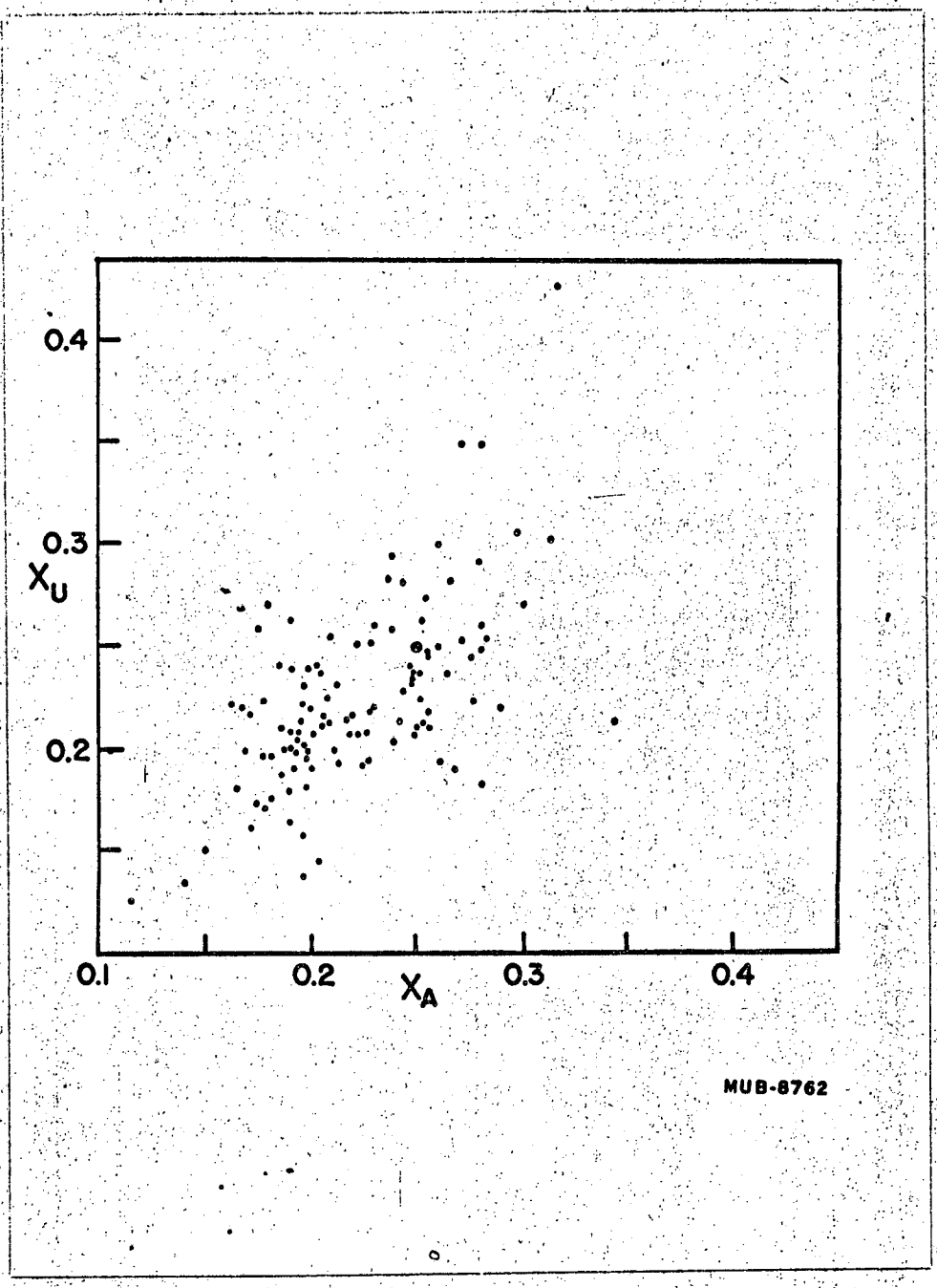


Figure 25. The base composition of various RNA's. A plot of the mole fraction of adenosine against the mole fraction of uridine for the 112 RNA's listed in Table VII.

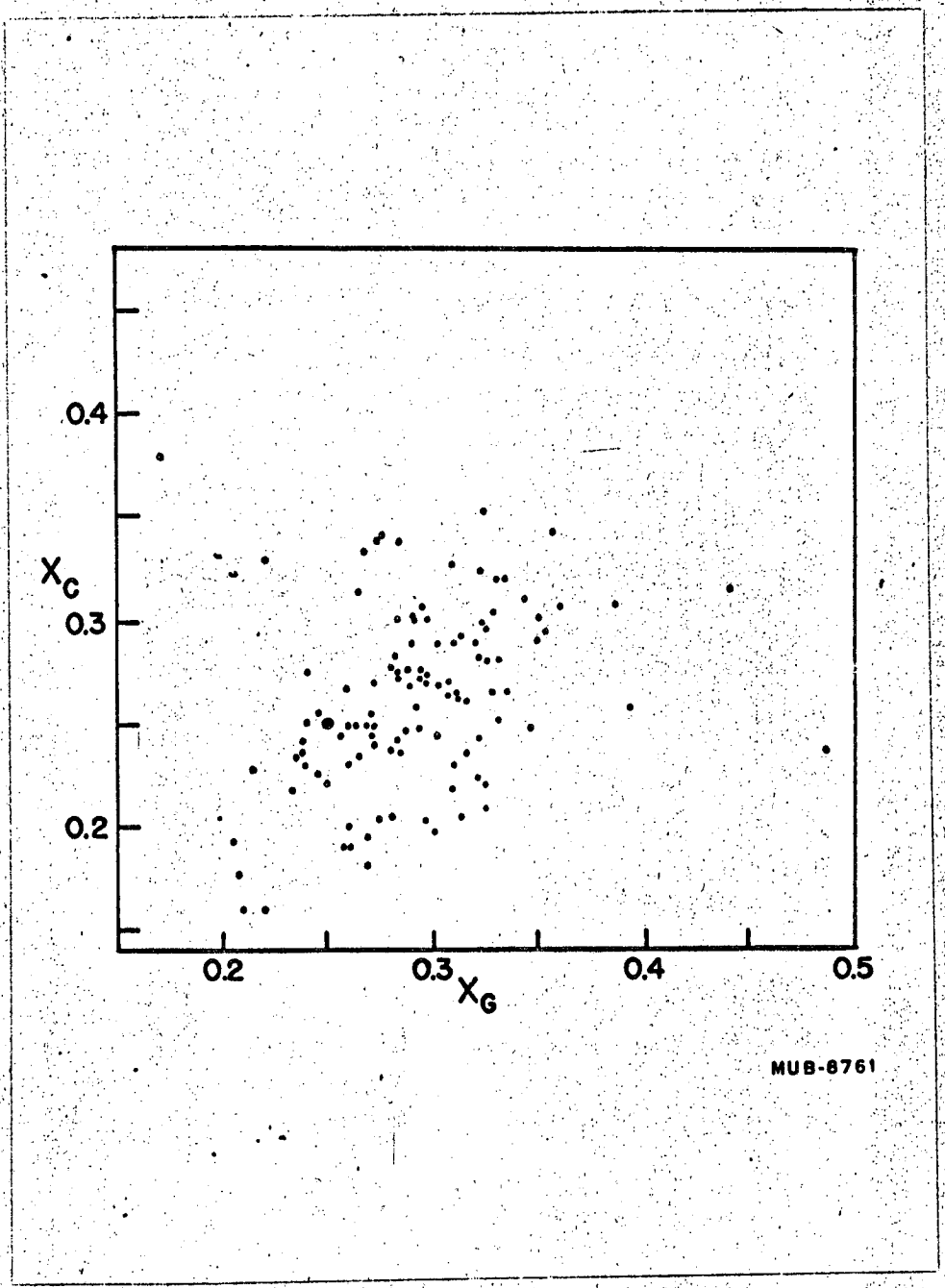


Figure 26. The base composition of various RNA's. A plot of the mole fraction of guanosine against the mole fraction of cytidine for the 112 RNA's listed in Table VII.

It indicates that there is more than one parameter governing the mole fraction of nucleotide in an RNA. Similar plots for DNA would show all of the points formed into fairly neat lines. If the base composition of RNA is purely dictated by chance, then the data should correlate in the following way. Let P_a represent the probability of having an A. $1-P_a$ is the probability of having anything other than A. Since there are three possibilities other than A, each of these should be equally probable given any value of A. Thus we would expect that $P_b = (1/3)(1-P_a)$. This means that plots of mole fraction A versus mole fraction B should have points clustered along lines of negative slope. This is what we observe in Figures 23 and 24 for plots of X_A versus X_C and X_G versus X_U. The scatter is not overly bad. By comparison, the plots of X_A versus X_U, and X_G versus X_C shown in Figures 25 and 26 show much worse scatter. But there is a definite tendency for the points to lie in a distribution with positive slope. This is just the opposite of what one expects for random statistics. Thus the more A an RNA has the more U it tends to have. The same is true for G and C. Return to Figures 23 and 24 and note that the best lines one could draw through the distribution of points in these two graphs would have a slope of -1. But the above discussion of random probability predicts that the slope should be -1/3. Thus there are definite deviations from random base composition in RNA. Similar results have been noticed previously by Elson and Chargaff, but they attributed this effect as arising from the impurity of the RNA's in their sample.[189] Most of the RNA data in Figures 23 through 26 comes from purified RNA. Thus this explanation is no longer valid.

We think the evidence shown in Figures 23 through 26 means that RNA's have a definite tendency to have DNA like base composition. If RNA is synthesized from both strands of DNA, this might be easy to reconcile. But many experiments have led to the idea that only one strand of RNA is copied in vivo. Bautz and Hall showed that the base composition of T4 messenger RNA is not like the DNA.[7] Later work by Bautz and Heding suggest that RNA is copied from DNA in an antiparallel fashion.[8] Finally, work described by Watson indicates that each and every RNA is copied from the same strand of DNA.[255] These elegant experiments were performed in the following way. The two complementary strands of SP8 virus can be easily separated since they have very different base composition. Each strand of DNA is annealed spearately with RNA synthesized by this phage. Complexes between RNA and DNA are observed to form only with one of the two DNA strands. These complexes can only occur when much of the sequence of the RNA is complementary to the DNA. Therefore, SP8 RNA is synthesized from one strand of the DNA.[255]

A second possible explanation of the above base composition data is that RNA (or, at least most RNA) is intentionally designed to be able to form many intramolecular hydrogen bonds. The number of canonical A-U and G-C hydrogen bonds which can be formed will approach a maximum value when A=U and G=C. Why should RNA tend to form double stranded segments? There appear to be two possibilities. RNA may prefer to be partially double stranded to facilitate carrying out its biological role. For example, it might be efficient for sRNA to have a single stranded anti codon but to have most of the remaining bases double stranded. This will decrease the possibility of copying errors.[56] Similarly, the ribosome

may be designed such that rRNA is mostly double stranded, except near portions of the polynucleotide strand which serve to bind the messenger or sRNA. A second possibility is simply that it is more efficient for RNA to have many double stranded sections. There are nucleases present in most cells, and a certain percentage of the RNA which is synthesized probably falls prey to them before it can serve any useful function. Most nucleases have a much higher activity towards single strands, and thus double stranded RNA is much less likely to be degraded by nucleases. Since evolution usually favors efficiency, it is possible that genes have been selected which produce partially RNAase-resistant RNA. However, we must not carry this argument too far, since a perfect double strand RNA is probably useless as a messenger. For whatever reasons, RNA's have apparently been designed to have more than an average number of A-U and G-C pairs.

14. Evidence for the Conformation of the Alanine sRNA

The last question to which we shall address ourselves is the conformation of the yeast alanine sRNA. This is the only RNA of known sequence, and it is reasonably small, containing only 77 nucleotides.[97] Thus it is a logical candidate for applying the methods we have developed in this chapter. Before we begin speculating on how the alanine sRNA is coiled up to form its conformation, we would do well to ask just how certain is the sequence determined by Holley and his coworkers. For among all of the sequence dependent properties we have mentioned, none is more sensitive than the double strands which an RNA might form. Recent work by Shapiro and his coworkers questions the validity of Holley's proposed

sequence.[202] They point out that there are several discrepancies between Holley's sequence and the experimental data. In fact, there are several sequences which fit more of Holley's data than his proposed sequence if all of the data is considered equivalent. But Holley has replied that some of his data are more conclusive than other of his experiments.[96] And we are inclined to agree with Holley's conclusions. Even if Holley were wrong and one of the sequences proposed by Shapiro et al. were right, this would not greatly affect the following discussion. Most of the proposals for correcting Holley's sequence involve removing a C from one part of the chain and replacing it with a G somewhere else. It turns out that the location of the residues in question prevents them from having any major effects on any of the probable conformations.

Before we discuss some of the detailed proposals for the conformation of the yeast alanine sRNA, we shall review some of the experimental data on the conformation of sRNA's. Most studies of sRNA physical properties have employed mixed yeast sRNA rather than a purified component. Thus we must be sure that all sRNA's are very similar if we are to apply data on mixtures to the pure alanine sRNA. Lindahl, Henley, and Fresco have shown that the molecular weight distribution of mixed yeast RNA is very narrow.[128] The number average molecular weight obtained from osmotic pressure data is 26,500 $\pm$ 300. The weight average molecular weight from sedimentation equilibrium is 26,150 $\pm$ 300. The molecular weight of pure yeast alanine sRNA calculated from the base composition is 26,600. All of these three values are in excellent agreement. Therefore, it appears that all sRNA's have almost exactly the same molecular weight. The range of chain lengths determined by Lindahl, Henley and Fresco from

analysis of the sedimentation equilibrium of the mixture is from 74 to 78.[128] The base composition of many of the purified sRNA's is very similar. Most of them have approximately as much A as U, and as much G as C. In most cases the ratio of A+U to G+C is about 2:3, (see Table VII). Thus we can conclude that there are great similarities among the components in mixed yeast sRNA.

From their study of hydrogen exchange in mixed sRNA, Englander and Englander conclude that there are 70 hydrogen bonded hydrogens per 70 nucleotides.[53] (They were apparently under the impression that sRNA had this degree of polymerization.) If we correct this value in the light of the more recent data by Lindahl, Henley and Fresco, we find that an average of 77 hydrogens are expected to be hydrogen bonded in sRNA. Englander and Englander estimate that this is 89% of the hydrogens available for hydrogen bonding. Thus sRNA is a highly helical molecule. Englander and Englander interpret their results as suggesting that most of the unusual and methylated bases found in sRNA are on single strand loops. They cite evidence that methylation is not necessary for biological competence <u>in vitro</u>.[53] This may mean that the methylating enzymes recognize their specific substrates by the fact that they are located in single stranded regions of the chain. This would suggest that the conformation of sRNA does not strongly depend on methylation.

Nihei and Cantoni have studied the digestion of sRNA by venom phosphodiesterase.[162] They find that a 40% digestion leads to a loss of 90% of the hypochromicity of the polymer. This evidence is very hard to reconcile with any of the structures we will discuss later. It is also in disagreement with our results on the hypochromicity of trinucleoside

diphosphates. For even if partial enzymatic digestion did break all of the double strand sections, the remaining single strands should still show a hypochromicity of around 20 to 30%. Since the hypochromicity of an sRNA is less than 50%, it is hard to see how half of an sRNA could have a hypochromicity of less than 10 to 15%. But the results of Nihei and Cantoni indicate that only 40% digestion reduces the hypochromicity to 5%. If their results are true, there must be serious complications which this argument ignores.

Fresco has reported the melting curves of several purified sRNA's.[66] He finds that the alanine sRNA has a biphasic melting curve in 0.15 M Na^+ at pH 7. The two transitions have T_m's of around 40°C and greater than 80°C. There is still some temperature dependence below 20°C. Valine and tyrosine sRNA show similar melting curves. But their extinction at 280 mμ reaches a steady plateau as the solution is cooled to room temperature. Felsenfeld and Cantoni have performed a detailed analysis of the temperature dependence of the UV spectrum of the serine sRNA.[56] The methods they used have been described earlier in this chapter. They predict three double strand regions in the serine sRNA. In order of stability these contain 4 A-U and 2 G-C pairs, 7 A-U and 7 G-C pairs, and no A-U and 7 G-C pairs. Unfortunately, the base sequence of the serine sRNA is not yet known, so these results cannot yet be used to construct a detailed model of the conformation. We would hesitate in applying these data directly to the alanine sRNA, since it would assume an unwarranted similarity among the structures of the sRNA's.

Fasman, Lindblow, and Seaman have measured the temperature transition of mixed yeast sRNA by both ORD and UV spectra.[55] They find that

the two melting curves do not coincide. The extinction at 260 mμ shows no changes at temperatures below about 50°C, while the ORD at the trough shows a marked dependence on temperature in this range. One possibility is that this simply represents the flexibility of the secondary structure. Glaubiger has shown that oscillations of neighboring bases can produce large changes in ORD and only small changes in the absorption spectrum.[74] But in the case of sRNA, it is possible that the difference in melting curves between UV and ORD measurements may represent a complex reshuffling of conformations.

Sarin and Zamecnik have noted that the ORD of sRNA changes when a polypeptide is attached to the CpCpA end.[194] They find that the amplitude of the long wavelength Cotton effect in an sRNA which is charged with 21 amino acids is 82% that of the free RNA. This change in conformation may well have biological importance. It is interesting that this change in conformation does not alter the UV spectrum. Thus it may represent a very subtle alteration.

The very recent hydrodynamic data of Henley, Lindahl, and Fresco offer some additional insights into the conformation of mixed yeast sRNA.[92] They compute that the axial ratio of sRNA lies between 4 and 7, though they admit that this value is not very accurate. Henley, Lindahl and Fresco report that the temperature dependence of the conformation sRNA occurs in two phases. The first, at temperatures less than 40°C, contains a substantial conformational change as measured by viscometry and sedimentation, but almost no change in secondary structure as measured by hypochromicity. At temperatures above 40°C, both the hydrodynamic and optical properties change together. This suggests that

the low temperature transition represents loss of tertiary structure with little breakage of double helical segments. At higher temperatures the double helices come apart and the bases in the single strands unstack.

The last experimental data on the alanine sRNA comes from the unpublished ORD results of Vournakis and Scheraga.[247] These were kindly made available to us. They show that the ORD of alanine sRNA is almost the same with or without the presence of Mg^{++}. The changes in the ORD upon heating are quite reversible. In addition, the shape of the ORD curve suggests that the RNA contains a large percentage of double strand. This will be discussed in more detail in the next section.

15. Models for the Conformation of the Alanine sRNA

Along with their published sequence of the yeast alanine sRNA, Holley and his coworkers presented several proposals for the conformation of this molecule in solution.[97] These were presumably devised with the intent to maximize the number of A-U and G-C pairs consistent with the postulates of Fresco, Alberts and Doty.[67] These are shown in Figure 27. In an attempt to find more possible conformations, we constructed a three dimensional model of the alanine sRNA. We wanted to take into account both the stacking of bases on a single strand and the possibility of base pairing. Thus we used strong springs to connect bases which we believe are stacked, and interlocking rings to connect unstacked bases. The latter permit relatively free rotation of linked residues. The former will attempt to keep stacked sections rigid. For the bases we used metal blocks of different sizes to account for the

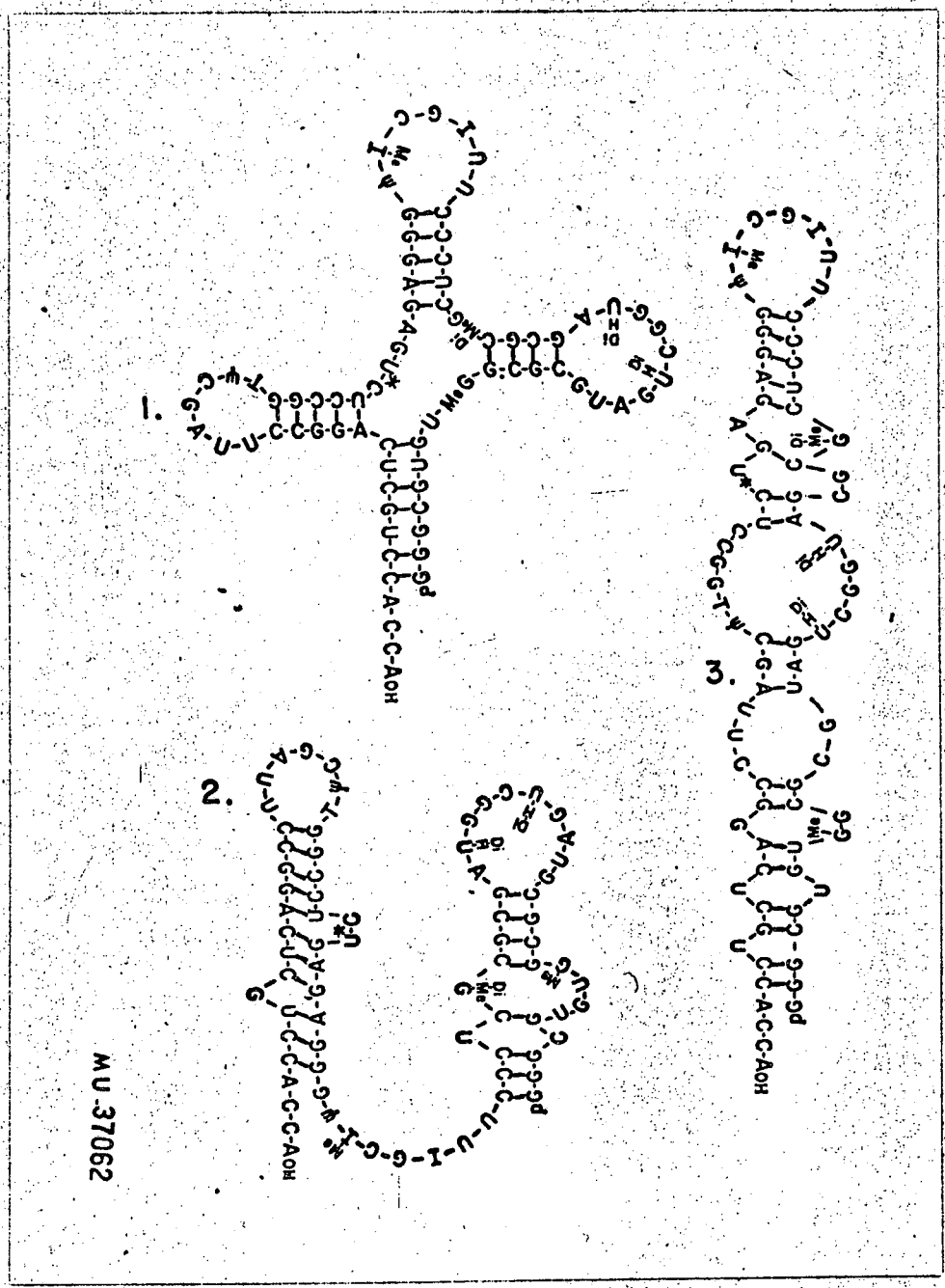
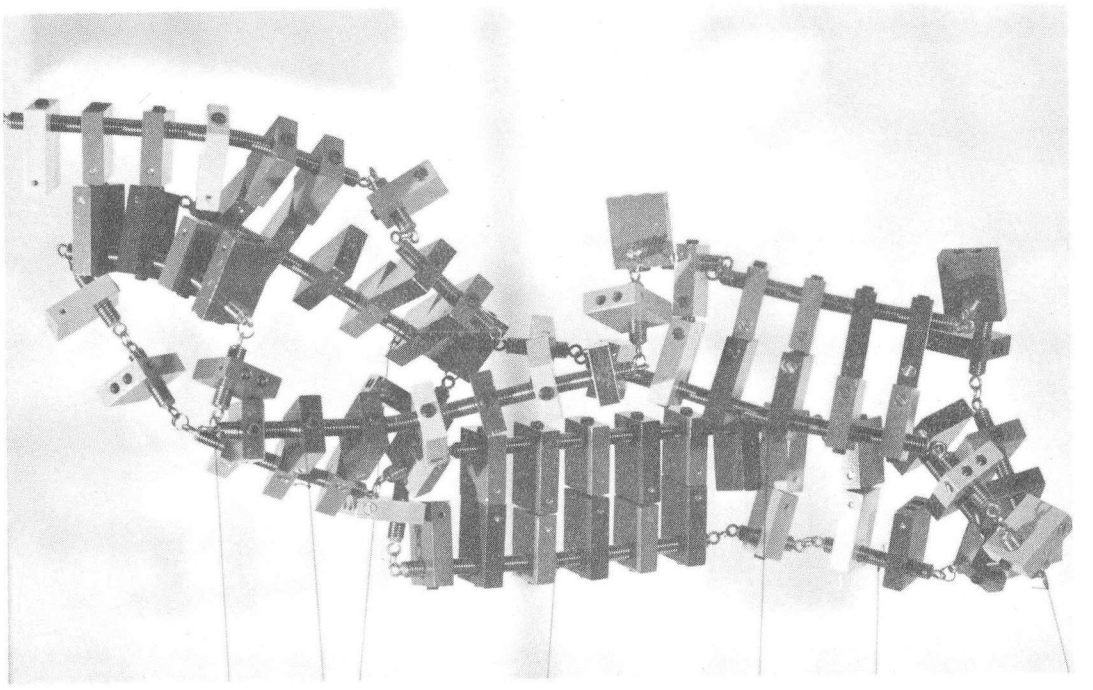


Figure 27. Several models for the conformation of the alanine sRNA which have been proposed by Holley and his collaborators. Adapted from Holley et al.[97]

larger volume of purines. This is an approximate scale model. With this model we were freed from the constraints of having to think in two dimensions. Thus we were able to find that it is possible to construct conformations for the alanine sRNA which have considerably more hydrogen bonds than the models proposed by Holley. At the same time, the construction of the model automatically took the additional constraints of base stacking into account. In Figure 28 is shown a photograph of the model we have built in the conformation which contains the maximum possible number of hydrogen bonded pairs consistent with stacking. The color code is as follows: G-blue, C-green, A-yellow, U-orange, and any other bases-red.

The following rules were used to select this conformation. We assumed that U does not stack as strongly as the other bases, and thus connected all U's by rings instead of springs. In addition, we assumed that the stacking of bases similar to U (T, dihydro U, and ψ) is also small. We assumed that methylated purines and I all stack strongly. We were also willing to consider hydrogen bond pairs in which one base was I, T or dihydro U. And we did not exclude the possibility of methylated bases being able to form hydrogen bonds, although we would expect these to be weaker than base pairs containing normal bases. The model shown in Figure 28 has several interesting features. Note that almost all loops or bends in the polynucleoside strand occur at U, T, dihydro U or ψ. In the model shown in Figure 28 we have violated one of the postulates described by Fresco, Alberts and Doty, since we form several hydrogen bonded pairs between sections of the chain which are far apart. But observation of the three dimensional models shows that these resi-



ZN-5373

Fig. 28. Photograph of a model of the alanine sRNA which contains 26 base pairs. The details of this model are discussed in the text.

dues are not far apart once some of the other base pairs in the structure have been formed. In an attempt to draw our model in two dimensions, Jukes has uncoupled these base pairs. He also wanted to free a possible anticodon, which we had formed into a double strand. His variation on our model is shown in Figure 29.[107] To change this drawing in our three dimensional model, one should form base pairs between residues 19 and 58, 20 and 57, 37 and 76, and 38 and 75, and let residue 47 bond to 67 instead of 8. (We begin counting from the 5' end of the chain.) We suspect that the first four base pairs are probably the weakest in the molecule, and probably come apart at about room temperature. This type of unfolding is consistent with the results observed for mixed sRNA by Henley, Lindahl and Fresco.[92] The axial ratio for the model shown in Figure 28 ranges from 3.7 to 7.1, depending on how closely the double strand segments are allowed to approach. This is in excellent agreement with the hydrodynamic data.[92]

A summary of various properties of the five models we have discussed for the alanine sRNA is given in Table VIII. It can be seen that our model shown in Figure 28 has more hydrogen bonded pairs than any other model. We form hydrogen bonds containing a total of 71 hydrogens. This is in surprisingly good agreement with the value of 77 which would be predicted by Englander and Englander for an sRNA containing 77 nucleotides. Thus our model is in fairly good agreement with almost all of the data presented thus far.

The one method we have not yet exploited is to compare the ORD of the alanine sRNA with values predicted for various conformations. In order to do this we must first compute the ORD of the single strand.

TABLE VIII

Models for the Conformation of the Yeast Alanine s-RNA

Model	Source	A-U Pairs	G-C Pairs	Hydrogen bonded hydrogens	Longest continuous double strand	Next longest continuous double strand
H-1 (Fig. 27-1)	Ref. 97	2	17	55	5 residues (2)	4 residues
H-2 (Fig. 27-2)	" 97	3	16	54	5 residues	4 residues
H-3 (Fig. 27-3)	" 97	4	14	50	5 residues	2 residues
J (Fig. 29)	" 107	5*	17†	61	5 residues (2)	4 residues
C (Fig. 28)	This work	5*	21††	71	5 residues (2)	4 residues

*Including one base pair between dihydro U and A.

†Including one base pair between dimethyl G and C (count as two hydrogen bonds).

††Including one base pair between I and C (count as two hydrogen bonds).

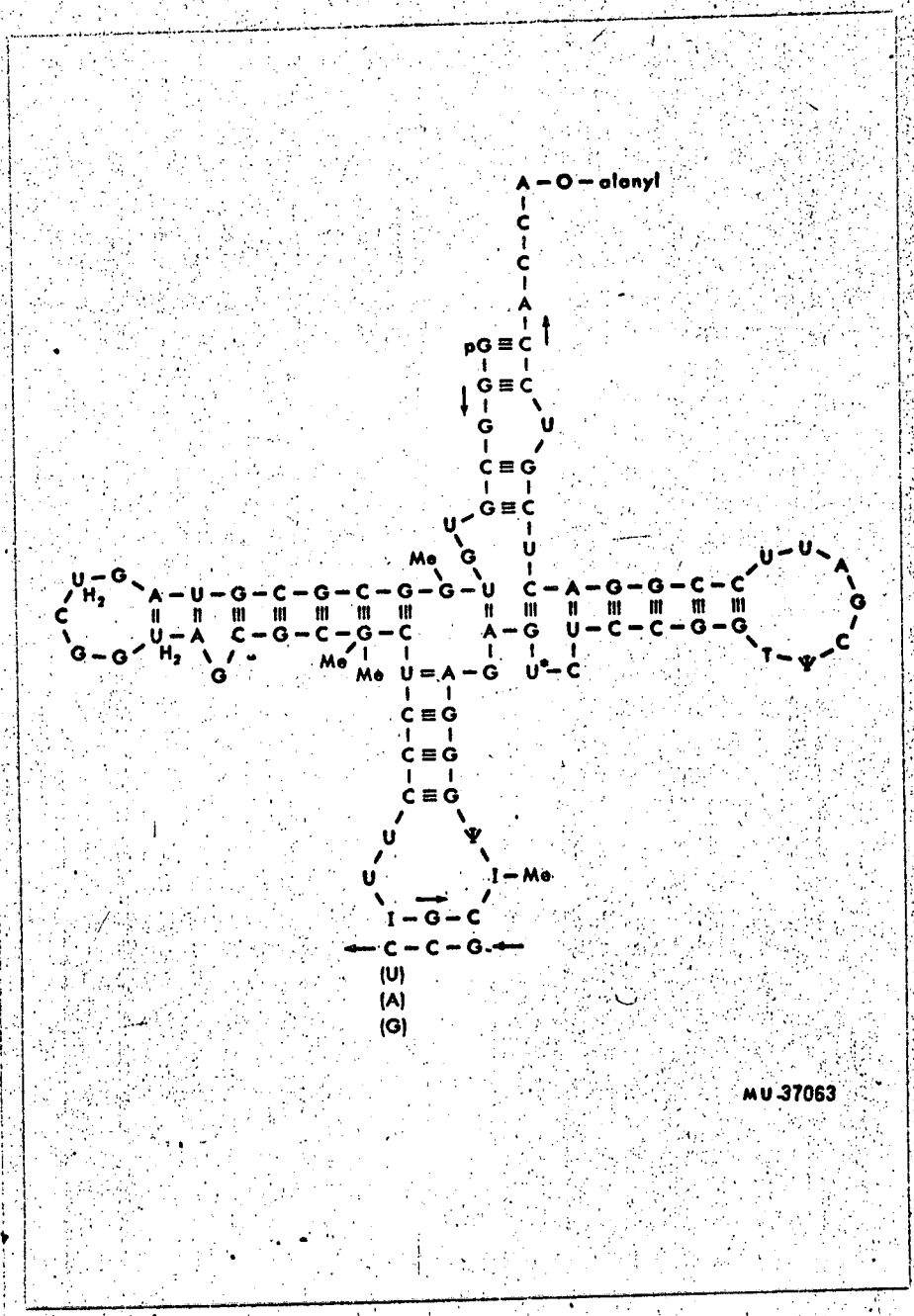


Figure 29. A suggestion for the conformation of the alanine sRNA which was adapted from the model shown in Figure 28 by Jukes.[107] Several possible codons are shown binding to a proposed anticodon. The details of this model are discussed in the text.

This was done using the methods we have described earlier in this chapter. The following approximations had to be made. We assumed that the contribution of unusual bases was the same as the normal bases they resemble. Thus we replaced I, MeI, MeG and DiMeG by G, ψ and T by U. Since dihydro U does not contain a chromaphore which absorbs in the 260 mμ region of the spectrum, we assumed it was a blank. Thus, optically, we treat the alanine sRNA as a mixture of three single strands. We expect that these approximations are not very severe, but until the ORD of oligomers containing strange bases is measured we will have no assurance. The ORD calculated for single strand alanine sRNA is compared with the experimental curve of Vournakis and Scheraga in Figure 30.[247] In this case, as in TMV RNA, the experimental curve (pH 6.83, 0.1 M phosphate, 0.15 M KCl) is shifted far to the blue of the values calculated for a single strand.

Since we have definite ideas for the conformation of the alanine sRNA we can use the methods described earlier to make an estimate of the effects of hydrogen bonding on the ORD of the RNA. To calculate the ORD difference curve for forming double strands from a single strand RNA, we need to know the fraction of residues forming A-U and G-C pairs. In the case of TMV RNA, we could make no detailed guess. But here we have five possible guesses. We have calculated the ORD difference curves for each of the five models using the equation below, where X is the mole fraction of base paired residue.

$$\Delta\phi = X_{A\text{-}U}\Delta\phi_{A\text{-}U} + X_{G\text{-}C}\Delta\phi_{G\text{-}C}$$

These results are shown in Figure 31. They are very encouraging since

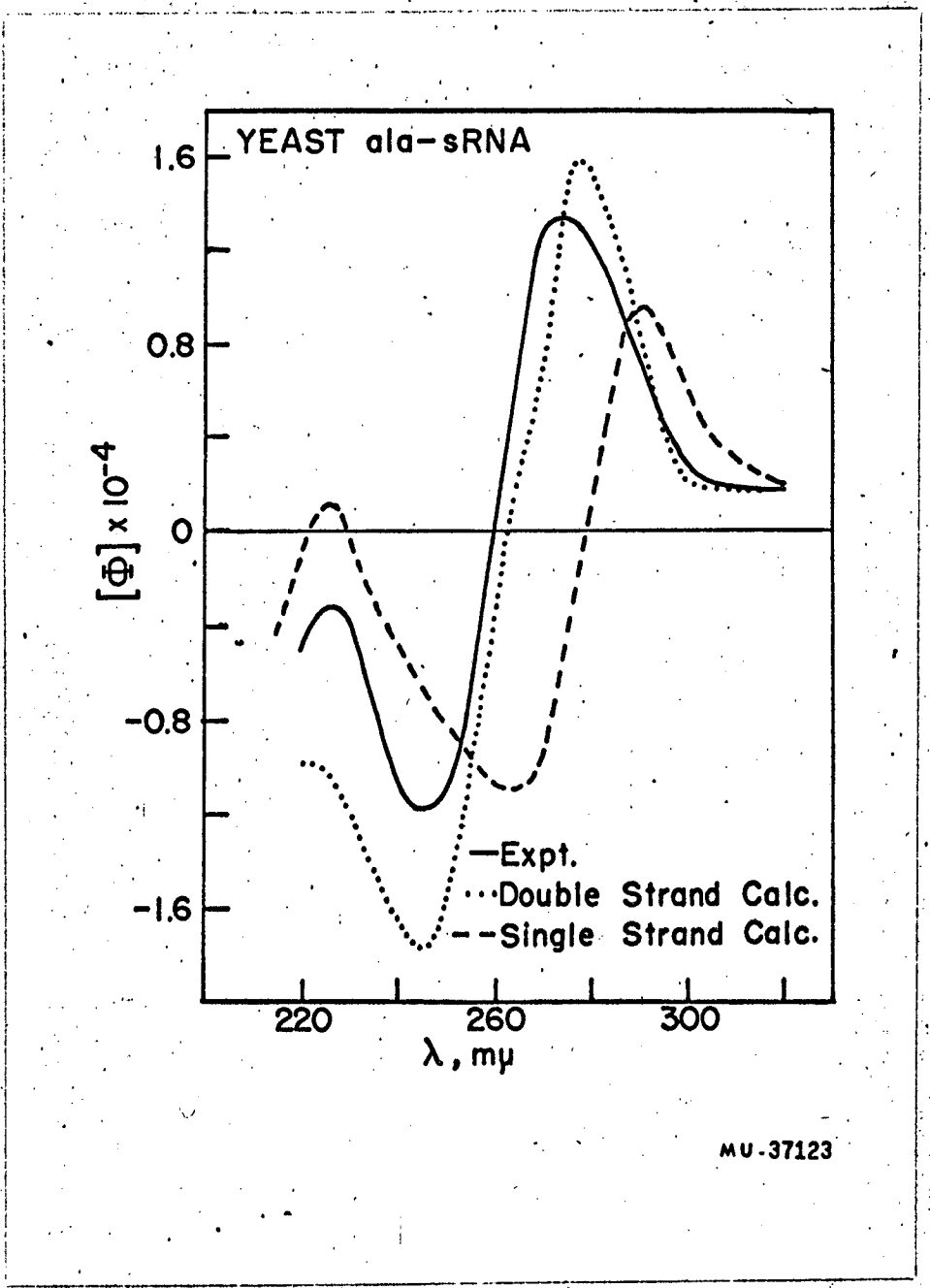


Figure 30. Optical rotatory dispersion of the yeast alanine sRNA.
——— experimental results of Vournakis and Scheraga[247]
– – – – – calculated ORD of single strand RNA
. . . . . calculated ORD of the conformation shown in Figure 28.

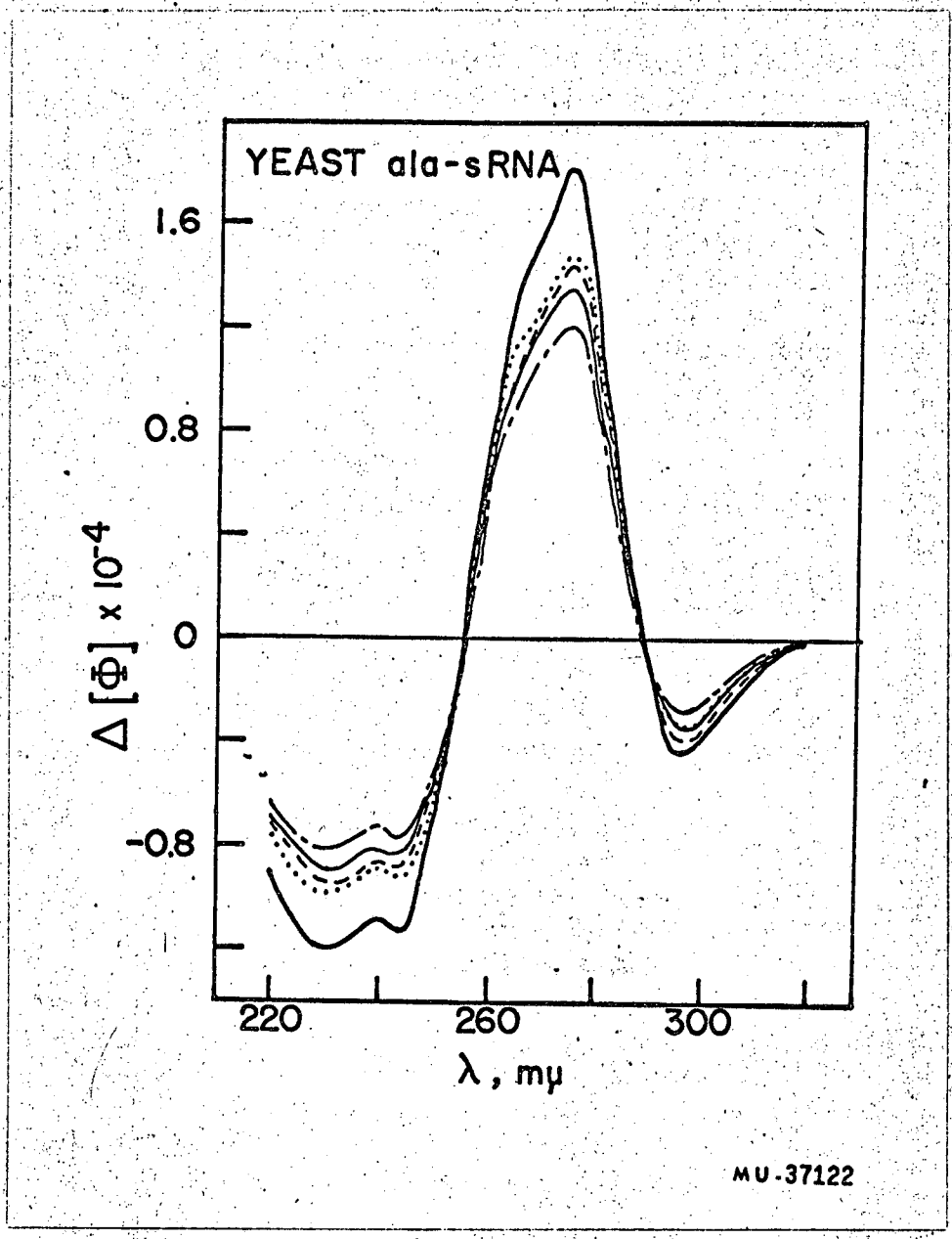


Figure 31. Calculated change in the ORD of the alanine sRNA upon forming a double strand conformation from a single strand. Results are shown for the five models summarized in Table VIII.

– – – – – Model H-1

———— Model H-2

— – — – Model H-3

• • • • • Model J

———— Model C

they suggest that small differences in the number and type of base pairs may be observable by measuring the ORD. In addition, we have not taken into account that for some of Holley's models there must be a decrease in stacking in order to form short loops in the middle of double strands. This will provide extra contributions to the ORD difference curves.

Finally, in Figure 30 we have included the calculated ORD of double stranded alanine sRNA, if it has the conformation shown in Figure 28. This was computed by adding the ORD difference curve to the calculated curve for the single stranded polymer. The agreement in the shapes of the experimental and calculated double strand ORD curves is excellent. By adding in the contributions of base pairs we are able to shift the single strand calculations until they have almost the same peak, trough, and cross-over as the experimental ORD. The magnitude of the calculated curve is less satisfactory since it is larger than the observed ORD. It is possible that this may be due to the choice of extinction coefficient. But a much more likely explanation is that the methods we have used here are very approximate. There are undoubtably errors in our calculation of the ORD of the double strand, since we have used only two of the ten contributing interactions. In addition, our model for the conformation may be far from the truth. The assumptions we made to count the contribution of unusual bases are also open to question. Thus, we are pleased that the agreement between calculation and experiment is as good as it is. Whether or not it is fortuitous remains to be seen. But we have shown in this chapter that ORD is potentially a very powerful tool for studying the intimate details of the conformation of RNA in solution.

ACKNOWLEDGEMENTS

I wish to express my sincere thanks to Professor Ignacio Tinoco, Jr. for awakening my interest in nucleic acids. His guidance in research has been invaluable, and I am especially grateful for many stimulating discussions, both scientific and nontechnical.

I have benefited greatly from fruitful collaboration with Dr. Leonard Peller, Dr. Stanley Mandeles, and Mr. Richard Jaskunas. This work has been aided by many interesting and provocative discussions with my fellow graduate students and several postdoctoral fellows. Dr. Alan Bush, Mr. Kenneth Cantor, Mr. Robert Davis, Dr. Steven Davis, Dr. Daniel Glaubiger, Mr. David Lloyd, Mr. Jean Thiery, Mr. Barrett Tomlinson, Dr. Myron Warshaw, and Dr. Rodney Biltonen, Dr. David McMullen, and Dr. Marcos Maestre have all been of great help. I acknowledge their many useful suggestions. I am indebted to many of the above, especially Dr. Myron Warshaw, and to Dr. H. Scheraga and Dr. J. T. Yang, for making unpublished results available to me.

I am grateful to Dr. Myron Warshaw for gifts of nucleotide material, and to Dr. Stanley Mandeles for frequent gifts of chemicals and TMV RNA. Miss Susan Dreyfuss has provided excellent technical assistance. I am especially thankful for her help in the preparation of polynucleotide phosphorylase.

I am greatly indebted to Mrs. Evie Litton for preparing all but a handful of the illustrations in this thesis. In addition, she has prepared most of the final manuscript. I am grateful for the patience of

Mrs. Johanna Onffroy, who typed most of the tables.

Finally, I would like to acknowledge my wife, Linda, for her patient understanding, for expert technical assistance, and for great help in preparing this manuscript.

I am grateful for partial support through a Cooperative Graduate Fellowship from the National Science Foundation. Additional support, and warm hospitality, have been received from the Laboratory of Chemical Biodynamics of the Lawrence Radiation Laboratory.

APPENDIX 1. FORTRAN PROGRAMS TO CALCULATE KINETICS OF EXONUCLEASE ACTION

These two programs calculate the kinetics of the liberation of monomeric units by exoenzyme degradation of a polymer. Both programs compute results for the random model of enzyme attack. The first, called 'exonuclease action', uses equation (12) of Chapter II. The second, called 'exonuclease action:steady state', employs equation (19) of Chapter II. It is shown in the text that these results can also be used for the optimal case of the random model, and for the two cases of the tight binding model.

The input for both programs should include the three rate constants, $k = k_1E$, k_2, and k_3, the initial concentration of enzyme substrate complex (usually set equal to zero), and the range of chain lengths covered by the degradation. In addition, the desired time increments and assymptotic values of the concentrations of each base must be specified.

The listing of the two programs is shown of the following pages, along with a sample of the output format.

```
      EXONUCLEASE ACTION
     ODIMENSION R(10),R2(10),R3(10),XO(10),DLFACT(252),BMAX(10),TINCR(10
     1),NMIN(10),NMAX(10)
      PRINT 102
  102 FORMAT (29H            EXONUCLEASE ACTION///)
C     CALCULATION OF LOG FACTORIAL
      TLOGX = 0.0
      DLFACT(1) = 0.0
      DO 11 K = 2,252
      AK = K-1
      PLOGX =LOGF(AK)
      TLOGX = TLOGX + PLOGX
      DLFACT(K) = TLOGX
   11 CONTINUE
      READ 104,NR
  104 FORMAT (I2)
      DO 400 J=1,NR
      READ 100,R(J),R2(J),R3(J),XO(J),BMAX(J),TINCR(J),NMIN(J),NMAX(J)
  100 FORMAT (6F10.0,2I5)
      PRINT 105
  105 FORMAT  (85H    K = K1E            K2            K3            X1(C)
     1      B(T) MAX        DELTA T   ///)
      PRINT 106,R(J),R2(J),R3(J),XO(J),BMAX(J),TINCR(J)
  106 FORMAT (1H 1P6E14.5)
      R=R(J)
      R2=R2(J)
      R3=R3(J)
      XO=XO(J)
      BMAX=BMAX(J)
      TINCR=TINCR(J)
      NMIN=NMIN(J)
      NMAX=NMAX(J)
      RR=0.5*SQRTF(R**2+R2**2+R3**2+2.*R*R2-2.*R*R3+2.*R2*R3)
      A1=-0.5*(R+R2+R3)+RR
      A2=-0.5*(R+R2+R3)-RR
      DO 400 N=NMIN,NMAX
      PRINT 110,N
  110 FORMAT (4HON= I4)
      PRINT 119
  119 FORMAT (21H    TIME        BASE  )
      T=0.000001
  300 E1=EXPF(A1*T)
      E2=EXPF(A2*T)
      SUM1 = 0.0
      SUM2=1.0
      SGN1=1.
      SGN2=(-1.)**N
      AN=N
      NN=N-1
      DO 200 I=0,NN
      AI=I
      NIP=N+I
      NIM=N-I
     0SUM1=SUM1+SGN1*EXPF(DLFACT(NIP)-DLFACT(N)-DLFACT(NIM)-DLFACT(I+1)
```

```
     1+(AN-1.-AI)*LOGF(T)-(AN+AI)*LOGF(2.*RR)+LOGF(E1+SGN2*E2)
     2+AN*LOGF(R*R3)-LOGF(R))
      SUMS =C.0
      SGN3=(-1.)**(N+I)
      DO 250 M=0,I
      AM=M
      NIMS=N+I-M
      IM=I-M
      EA=E1*(R**N/A1**(M+1))+SGN3*E2*(R**N/A2**(M+1))
      IF(EA) 601,602,602
  601 EA=-EA
      S4=-1.
      GO TO 603
  602 S4=1.
  6030SUMS=SUMS+S4*EXPF(DLFACT(NIMS)-DLFACT(IM+1)-(AN+AI-AM)*LOGF(2.*RR)
     1+LOGF(EA))
  250 SGN3=-SGN3
      IF (SUMS) 604,605,605
  604 SUMS=-SUMS
      S5=-1.
      GO TO 666
  605 S5=1.
  6660SUM2=SUM2+SGN1*S5*EXPF(-DLFACT(N)-DLFACT(NIM)+(AN-AI-1.)*LOGF(T)+
     1LOGF(SUMS)+AN*LOGF(R3))
      SGN1=-SGN1
  200 SGN2=-SGN2
      B=XO*SUM1+SUM2
      PRINT 120,T,B
  120 FORMAT (1H F10.2, F10.4)
      IF(B-BMAX) 205,205,400
  205 T=T+TINCR
      GO TO 300
  400 CONTINUE
      CALL EXIT
      END(1,1,0,0,0,0,0,0,0,0,0,0,0,0,0)
```

```
      EXONUCLEASE ACTION: STEADY-STATE
     ODIMENSION R(10),R2(10),R3(10),XO(10),DLFACT(252),BMAX(10),TINCR(10
     1),NMIN(10),NMAX(10)
      PRINT 102
  102 FORMAT (29H          EXONUCLEASE ACTION///)
C     CALCULATION OF LOG FACTORIAL
      TLOGX = 0.0
      DLFACT(1) = 0.0
      DO 11 K = 2,252
      AK = K-1
      PLOGX =LOGF(AK)
      TLOGX = TLOGX + PLOGX
      DLFACT(K) = TLOGX
   11 CONTINUE
      READ 104,NR
  104 FORMAT (I2)
      DO 400 J=1,NR
      READ 100,R(J),R2(J),R3(J),XO(J),BMAX(J),TINCR(J),NMIN(J),NMAX(J)
  100 FORMAT (6F10.0,2I5)
      PRINT 105
 1050FORMAT  (85H    K = K1E            K2            K3            X1(C)
     1      B(T) MAX       DELTA T    //)
      PRINT 106,R(J),R2(J),R3(J),XO(J),BMAX(J),TINCR(J)
  106 FORMAT (1H 1P6E14.5)
      R=R(J)
      R2=R2(J)
      R3=R3(J)
      XO=XO(J)
      BMAX=BMAX(J)
      TINCR=TINCR(J)
      NMIN=NMIN(J)
      NMAX=NMAX(J)
      RR=0.5*SQRTF(R**2+R2**2+R3**2+2.*R*R2-2.*R*R3+2.*R2*R3)
      A1=-0.5*(R+R2+R3)+RR
      A2=-0.5*(R+R2+R3)-RR
      DO 400 N=NMIN,NMAX
      PRINT 110,N
  110 FORMAT (4HON= I4)
      PRINT 119
  119 FORMAT (21H   TIME       BASE  )
      T=0.000001
  300 CONTINUE
      SUM3=0.0
      NN=N-1
      DO 200 I=0,NN
      AI=I
      SUM3=SUM3+EXPF(AI*LOGF(T)-DLFACT(I+1))
  200 CONTINUE
      B=1.-EXPF(-T)*SUM3
      PRINT 120,T,B
  120 FORMAT (1H F10.2, F10.4)
      IF(B-BMAX) 205,205,400
  205 T=T+TINCR
      GO TO 300
  400 CONTINUE
      CALL EXIT
      END(1,1,0,0,0,0,0,0,0,0,0,0,0,0,0)
```

Sample Output from Either Program

K = K1E	K2	K3	X1(0)	B(T) MAX	DELTA T
1.00000E 03	1.00000E 00	10.00000E-04	0.	9.50000E-01	1.00000E 03

N= 1

TIME	BASE
0.00	-0.0000
1000.00	0.6316
2000.00	0.8641
3000.00	0.9497
4000.00	0.9812

N= 2

TIME	BASE
0.00	-0.0000
1000.00	0.2638
2000.00	0.5932

APPENDIX 2. FORTRAN PROGRAMS TO CALCULATE THE OPTICAL ROTATORY DISPERSION OF OLIGONUCLEOTIDES AND POLYNUCLEOTIDES

These two programs can calculate the molar rotation of either oligo- or polynucleotides. Both contain an algorithm which automatically selects whether to use equations similar to equations (5) or (12) of Chapter III, if the following instructions are obeyed. To specify an oligomer in either of these two programs, it is necessary to include as input the number of times each possible dimer sequence appears. Also needed is the number of times each monomer appears, except the two terminal residues. Thus, to specify ApGpC, one needs one AG, one GC, and one G. To specify a polymer, the number of times each sequence appears is replaced by the mole fraction of dimer in the polymer. Also needed to calculate polymer ORD is the mole fraction of each monomer. The dimer and monomer ORD data needed to perform these calculations are described in Chapter III. These results appear in a library at the beginning of the data deck of each program.

The first program can calculate the ORD of a single monomer or polymer. It automatically computes the molar rotation at 2.5 mμ intervals from 332.5 to 215.0 mμ. The second program can calculate the ORD of mixtures of oligomers or polymers which contain up to ten different components. Each component is specified by the rules described above. Additional input includes the composition of the mix-

ture, and the range of wavelengths and intervals desired. The range can be as large as from 332.5 mμ to 215 mμ, and the intervals can be any integral multiple of 2.5 mμ. Listings of these two programs and samples of their output format are shown on the next few pages.

Polymer ORD

```
C       PROGRAM TO CALCULATE POLYMER ORD
        DIMENSION FD(16),FM(4),PHID(100,16),PHIM(100,4),PHIP(100),
       1ALAM(100),S1(100),S2(100),SPOT(12)
C       DIMER AND MONOMER DATA FROM 215.0 TO 332.5 IN 2.5 INCREMENTS
        READ 5,((PHID(K,I),K=2,49),I=1,16)

     5  FORMAT (12F6.2)
        READ 6,((PHIM(K,J),K=2,49),J=1,4)

     6  FORMAT (12F6.2)
C       ORDER OF DIMERS AA AU AC AG UA UU UC UG CA CU CC CG GA GU GC GG
C       ORDER OF MONOMERS A U C G
    55  READ 8, (SPOT(M),M=1,12)
     8  FORMAT (12A6)
        PRINT 9, (SPOT(M),M=1,12)
     9  FORMAT (1H1,10X,12A6///)
        PRINT 99
    99  FORMAT (51X,18HDIMER  COMPOSITION//)
        PRINT 100
   100  FORMAT    (110H    AA       AU       AC       AG       UA       UU
       1UG      CA       CU       CC       CG       GA       GU       GC
       1       GG//)
        READ 18,FD
    18  FORMAT (8F7.3)
        PRINT 19,FD
    19  FORMAT (16F7.3///)
        PRINT 199
   199  FORMAT (50X,20HMONOMER  COMPOSITION//)
        PRINT 200
   200  FORMAT (42X,27H    A        U        C        G  //)
        READ 23, FM
    23  FORMAT (4F7.3)
        PRINT 24, FM
    24  FORMAT (42X,4F7.3///)
        ALAM(1)=335.0
        SUM1=0.0
        DO 101 I=1,16
   101  SUM1=SUM1+FD(I)
        SUM2=0.0
        DO 102 J=1,4
   102  SUM2=SUM2+FM(J)
        CNORM=2.*SUM1-SUM2
        DO 473 K=2,49
        S1(K)=0.0
        DO 15 I=1,16
    15  S1(K)=S1(K)+2.*FD(I)*PHID(K,I)
        S2(K)=0.0
        DO 16 J=1,4
    16  S2(K)=S2(K)+FM(J)*PHIM(K,J)
        PHIP(K)=(S1(K)-S2(K))/CNORM
   473  ALAM(K) = ALAM(K-1) -2.5
        PRINT 299
   299  FORMAT (34HLAMBDA   MOLAR ROTATION X 1/10,000//)
        PRINT 300,(ALAM(K),PHIP(K),K=2,49)
   300  FORMAT (F6.1,F16.3)
        GO TO 55
C       ORDER OF DATA IS PHID,PHIM,SPOT,FD,FM
        CALL EXIT
        STOP
        END
```

RANDOM EQUIMOLAR RNA

DIMER COMPOSITION

AA	AU	AC	AG	UA	UU	UC	UG
0.062	0.062	6.000	6.000	0.062	0.062	0.062	0.062

CA	CU	CC	CG	GA	GU	GC	GG
0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.062

MONOMER COMPOSITION

A	U	C	G
0.250	0.250	0.250	0.250

LAMBDA	MOLAR ROTATION X 1/10,000
332.5	0.044
330.0	0.054
327.5	0.064
325.0	0.074
322.5	0.084
320.0	0.090
317.5	0.095
315.0	0.110
312.5	0.126
310.0	0.141
307.5	0.172
305.0	0.202
302.5	0.243
300.0	0.329
297.5	0.400
295.0	0.491
292.5	0.554
290.0	0.596
287.5	0.578
285.0	0.478
282.5	0.299
280.0	0.097
277.5	-0.247
275.0	-0.687
272.5	-1.050
270.0	-1.281
267.5	-1.443
265.0	-1.451
262.5	-1.326
260.0	-1.156
257.5	-0.915

```
    Mixture   ORD
C       SYNTHESIS OF ORD OF MIXTURES
        DIMENSION PHID(100,16),PHIM(100,4),SPOT(12),FD(10,16),FM(10,4),KLA
       1M(100),CNORM(10),FMIX(10,10),S(100)
        READ 1,KDEL,KMIN,KMAX
      1 FORMAT (3I5)
        KM=(KMAX-KMIN)/KDEL+2
        DO 2 I=1,16
      2 READ 3,(PHID(K,I),K=2,KM)
      3 FORMAT (12F6.2)
        DO 4 I=1,4
      4 READ 5,(PHIM(K,I),K=2,KM)
      5 FORMAT (12F6.2)
     27 READ 6,(SPOT(M),M=1,12)
      6 FORMAT (12A6)
        PRINT 7,(SPOT(M),M=1,12)
      7 FORMAT (1H1,10X,12A6///)
        READ 8,NMIX,NCOM
      8 FORMAT (2I3)
        IF(NMIX) 100,100,101
    101 DO 9 J=1,NCOM
        READ 10,(FD(J,I),I=1,16)
     10 FORMAT (8F7.3)
      9 READ 11,(FM(J,I),I=1,4)
     11 FORMAT (4F7.3)
        KLAM(1)=KMAX+KDEL
        DO 12 N=1,NMIX
     12 READ 13,(FMIX(N,J),J=1,NCOM)
     13 FORMAT (10F7.3)
        DO 14 J=1,NCOM
        SUM1=0.0
        DO 15 I=1,16
     15 SUM1=SUM1+FD(J,I)
        SUM2=0.0
        DO 16 I=1,4
     16 SUM2=SUM2+FM(J,I)
     14 CNORM(J)=2.*SUM1-SUM2
        DO 17 N=1,NMIX
        DO 19 K=2,KM
        S1=0.0
        S2=0.0
        DO 22 J=1,NCOM
        DO 21 I=1,16
     21 S1=S1+2.*PHID(K,I)*FD(J,I)*FMIX(N,J)/CNORM(J)
        DO 22 I=1,4
     22 S2=S2+PHIM(K,I)*FM(J,I)*FMIX(N,J)/CNORM(J)
     20 S(K)=S1-S2
     19 KLAM(K)=KLAM(K-1)-KDEL
        PRINT 23
     23 FORMAT (109H  AA   AU   AC   AG   UA   UU   UC   UG   CA   CU   C
       1C   CG   GA   GU   GC   GG   A    U    C    G   FRACTION/)
        PRINT 24,((FD(J,I),I=1,16),(FM(J,I),I=1,4),FMIX(N,J),J=1,NCOM)
     24 FORMAT (20F5.2,F7.3)
        PRINT 25
     25 FORMAT(34HLAMBDA   MOLAR ROTATION X 1/10,000/)
     17 PRINT 26,(KLAM(K),S(K),K=2,KM)
     26 FORMAT (I5,F16.3)
        GO TO 27
C       ORDER OF DATA=KDEL,KMIN,KMAX-PHID(16)-PHIM(4)-SPOT-NMIX,NCOM-FD(1
C       6-FM(4)-FMIX(NMIX)
C       TO END USE BLANK CARD PLUS 0 IN COLUMES 3 AND 6
    100 CALL EXIT
        STOP
        END
```

MIXTURE OF GAC AND AGC

AA	AU	AC	AG	UA	UU	UC	UG	CA	CU	CC	CG	GA	GU	GC	GG	A	U	C	G	FRACTION
0.	0.	0.	1.00	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	1.00	0.	0.	0.	0.	1.00	1.000
0.	0.	1.00	0.	0.	0.	0.	0.	0.	0.	0.	0.	1.00	0.	0.	0.	1.00	0.	0.	0.	0.

LAMBDA	MOLAR ROTATION X 1/10,000
3325	0.080
3300	0.093
3275	0.107
3250	0.117
3225	0.123
3200	0.130
3175	0.133
3150	0.140
3125	0.153
3100	0.170
3075	0.203
3050	0.250
3025	0.290
3000	0.380
2975	0.450
2950	0.540
2925	0.583
2900	0.560
2875	0.483
2850	0.320
2825	0.153
2800	0.003
2775	-0.263
2750	-0.547
2725	-0.700
2700	-0.840
2675	-0.857
2650	-0.777
2625	-0.640
2600	-0.587
2575	-0.430
2550	-0.227
2525	-0.160

APPENDIX 3. FORTRAN PROGRAM TO DETERMINE THE BASE COMPOSITION OF AN UNDEGRADED OLIGOMER

We are in the process of developing a method which can determine the base composition of an oligomer without having to degrade it to monomers. This approach is based on the assumption that there exist conditions under which the UV spectrum of the oligomer is the sum of the spectra of its monomer components. Such is the case in solutions of dimers and trimers in seven molar urea, at pH 1.5, 7, and 11.[143] Thus, we would like a simple way of comparing the spectra of a pure oligomer with the sum of the spectra of its monomers. The program reproduced in this section solves this problem in a very simple manner. It computes the UV spectrum of every possible base composition for a given chain length, and compares each of the spectra with experimental results. The comparison is done by computing the sum of the squares of the deviations at 96 wavelengths, which range from 325 mμ to 220 mμ in 1 mμ intervals. The resulting sums of deviations are grouped in increasing order, and appear as output along with their corresponding base composition. Thus, if one has a pure oligomer, it is possible to determine the base composition in this manner. A typical set of results is shown at the end of the program listing. From these results it can be seen that if unknown #2 is a pure dimer, it almost certainly contains one A and one G. In fact, unknown #2 turned out to be ApG.

The necessary input for this program consists of a library of UV spectra of A, U, C, and G at pH 1.5, 7, and 11. The range of wavelengths and wavelength increments can be adjusted. Spectra of oligomers, taken under the same conditions, provide the remainder of the input. The listing of this Fortran program is shown on the following pages.

```
C     COMPARISON OF UNKNOWN SPECTRUM WITH OLIGOMERS OF LENGTH NN
      DIMENSION FO(56,4),F(10,4),D(20,4),T3(35,4),T4(56,4),P(56,4),NFACT
     1(8),EXT1(140,4),EXT2(140,4),EXT3(140,4),UNK1(140),UNK2(140),UNK3(1
     240),SSQ1(56),SSQ2(56),SSQ3(56),SSQ(56),SPOT(12),RSQ(56)
      READ 1,KDEL,KMIN,KMAX
    1 FORMAT (3I5)
      KM=(KMAX-KMIN)/KDEL +1
      K26=(KMAX-2600)/KDEL +1
      READ 2,NMAX
    2 FORMAT (I2)
      NM=NMAX+3
      NFACT(1)=1
      DO 3 K=2,NM
    3 NFACT(K)=NFACT(K-1)*K
      DATA (F(1,J),J=1,4)/1.0,3*0./
      DATA (F(2,J),J=1,4)/0.,1.0,2*0./
      DATA (F(3,J),J=1,4)/2*0.,1.0,0./
      DATA (F(4,J),J=1,4)/3*0.,1.0/
      DO 11 I=1,4
      DO 11 J=1,4
   11 D(I,J)=F(1,J)+F(I,J)
      DO 13 I=5,7
      DO 13 J=1,4
   13 D(I,J)=F(2,J)+F(I-3,J)
      DO 15 I=8,9
      DO 15 J=1,4
   15 D(I,J)=F(3,J)+F(I-5,J)
      DO 17 J=1,4
   17 D(10,J)=F(4,J)+F(4,J)
      IF(NMAX-2) 42,42,18
   18 DO 19 I=1,10
      DO 19 J=1,4
   19 T3(I,J)= F(1,J)+D(I,J)
      DO 21 I=11,16
      DO 21 J=1,4
   21 T3(I,J)=F(2,J)+D(I-6,J)
      DO 23 I=17,19
      DO 23 J=1,4
   23 T3(I,J)=F(3,J)+D(I-9,J)
      DO 25 J=1,4
   25 T3(20,J)=F(4,J)+D(10,J)
      IF(NMAX-3) 99,42,26
   26 DO 27 I=1,20
      DO 27 J=1,4
   27 T4(I,J)=F(1,J)+T3(I,J)
      DO 29 I=21,30
      DO 29 J=1,4
   29 T4(I,J)=F(2,J)+T3(I-10,J)
      DO 31 I=31,34
      DO 31 J=1,4
   31 T4(I,J)=F(3,J)+T3(I-14,J)
```

```
      DO 33 J=1,4
   33 T4(35,J)=F(4,J)+T3(20,J)
      IF(NMAX-4) 99,42,34
   34 DO 35 I=1,35
      DO 35 J=1,4
   35 P(I,J)=F(1,J)+T4(I,J)
      DO 37 I=36,50
      DO 37 J=1,4
   37 P(I,J)=F(2,J)+T4(I-15,J)
      DO 39 I=51,55
      DO 39 J=1,4
   39 P(I,J)=F(3,J)+T4(I-20,J)
      DO 41 J=1,4
   41 P(56,J)=F(4,J)+T4(35,J)
   42 READ 58,((EXT1(K,J),K=1,KM),J=1,4)

   58 FORMAT (12F6.3)
      READ 60,((EXT2(K,J),K=1,KM),J=1,4)

   60 FORMAT (12F6.3)
      READ 62,((EXT3(K,J),K=1,KM),J=1,4)

   62 FORMAT (12F6.3)
  142 READ 43,NN,NOUT
   43 FORMAT (2I3)
      IF(NN) 99,99,44
   44 NNN=NFACT(NN+3)/(NFACT(NN)*NFACT(3))
  144 READ 45,(SPOT(I),I=1,12)
   45 FORMAT (12A6)
      READ 73,(UNK1(K),K=1,KM)
   73 FORMAT (12F6.3)
      READ 74,(UNK2(K),K=1,KM)
   74 FORMAT (12F6.3)
      READ 75,(UNK3(K),K=1,KM)
   75 FORMAT (12F6.3)
      UN1ST=UNK1(K26)
      UN2ST=UNK2(K26)
      UN3ST=UNK3(K26)
      DO 76 K=1,KM
      UNK1(K)=UNK1(K)/UN1ST
      UNK2(K)=UNK2(K)/UN2ST
   76 UNK3(K)=UNK3(K)/UN3ST
      GO TO (46,48,50,52,54),NN
   46 DO 47 I=1,NNN
      DO 47 J=1,4
   47 FO(I,J)=F(I,J)
      GO TO 56
   48 DO 49 I=1,NNN
      DO 49 J=1,4
   49 FO(I,J)=D(I,J)
      GO TO 56
   50 DO 51 I=1,NNN
```

```
      DO 51 J=1,4
   51 FO(I,J)=T3(I,J)
      GO TO 56
   52 DO 53 I=1,NNN
      DO 53 J=1,4
   53 FO(I,J)=T4(I,J)
      GO TO 56
   54 DO 55 I=1,NNN
      DO 55 J=1,4
   55 FO(I,J)=P(I,J)
   56 DO 79 I=1,NNN
      ABST1=0.0
      ABST2=0.0
      ABST3=0.0
      SSQ1(I)=0.0
      SSQ2(I)=0.0
      SSQ3(I)=0.0
      DO 111 J=1,4
      ABST1=ABST1+EXT1(K26,J)*FO(I,J)
      ABST2=ABST2+EXT2(K26,J)*FO(I,J)
  111 ABST3=ABST3+EXT3(K26,J)*FO(I,J)
      DO 77 K=1,KM
      ABS1=0.0
      ABS2=0.0
      ABS3=0.0
      DO 67 J=1,4
      ABS1=ABS1+EXT1(K,J)*FO(I,J)
      ABS2=ABS2+EXT2(K,J)*FO(I,J)
   67 ABS3=ABS3+EXT3(K,J)*FO(I,J)
      ABS1=ABS1/ABST1
      ABS2=ABS2/ABST2
      ABS3=ABS3/ABST3
      SSQ1(I)=SSQ1(I)+(UNK1(K)-ABS1)**2
      SSQ2(I)=SSQ2(I)+(UNK2(K)-ABS2)**2
   77 SSQ3(I)=SSQ3(I)+(UNK3(K)-ABS3)**2
      SSQ(I)=SSQ1(I)+SSQ2(I)+SSQ3(I)
   79 RSQ(I)=SSQ(I)**.5
      NM1=NNN-1
      DO 85 I=1,NM1
      IP1=I+1
      DO 85 K=IP1,NNN
      IF(SSQ(I)-SSQ(K)) 85,85,83
   83 TEMP=SSQ(I)
      SSQ(I)=SSQ(K)
      SSQ(K)=TEMP
      DO 82 J=1,4
      TEMP=FO(I,J)
      FO(I,J)=FO(K,J)
   82 FO(K,J)=TEMP
      TEMP=SSQ1(I)
      SSQ1(I)=SSQ1(K)
      SSQ1(K)=TEMP
```

```
      TEMP=SSQ2(I)
      SSQ2(I)=SSQ2(K)
      SSQ2(K)=TEMP
      TEMP=SSQ3(I)
      SSQ3(I)=SSQ3(K)
      SSQ3(K)=TEMP
      TEMP=RSQ(I)
      RSQ(I)=RSQ(K)
      RSQ(K)=TEMP
   85 CONTINUE
      PRINT 86,(SPOT(I),I=1,12)
   86 FORMAT (1H1,10X,12A6///)
      PRINT 88
   88 FORMAT (14HKMIN KMAX KDEL//)
      PRINT 90,KMIN,KMAX,KDEL
   90 FORMAT (3I5///)
      PRINT 87
   87 FORMAT    (69H      SSQ        RSQ      SSQPH1     SSQPH7     SSQPH11
     1A     U     C     G//)
  189 PRINT 89,(SSQ(I),RSQ(I),SSQ1(I),SSQ2(I),SSQ3(I),(FO(I,J),J=1,4),I=
     11,NOUT)
   89 FORMAT (5F10.5,4F5.0)
      GO TO 142
C     WAVELENGTH DATA FROM KMAX TO KMIN
C     ORDER OF DATA.NMAX,,KMIN,KMAX,KDEL,,EXT1,EXT2,EXT3,,NN,NOUT,,SPOT,
C     ,UNK1,UNK2,UNK3
   99 CALL EXIT
      STOP
      END
```

UNKNOWN 2 JUNE 16 AS DIMER

KMIN	KMAX	KDEL
2250	3200	10

SSQ	RSQ	SSQPH1	SSQPH7	SSQPH11	A	U	C	G
0.28299	0.53197	0.06604	0.09560	0.12135	1.	0.	0.	1.
0.99640	0.99820	0.25130	0.31624	0.42886	0.	1.	0.	1.
2.74479	1.65674	0.95298	1.32830	0.46350	1.	1.	0.	0.
3.11096	1.76379	1.11783	1.31011	0.68303	0.	2.	0.	0.
3.69037	1.92103	1.24128	1.34413	1.10497	0.	0.	0.	2.
3.86057	1.96483	1.01938	1.67488	1.16630	2.	0.	0.	0.
5.00200	2.23652	3.83648	0.77802	0.38751	1.	0.	1.	0.
17.38297	4.16929	9.76045	2.89168	4.73084	0.	1.	1.	0.
20.66744	4.54615	12.44121	3.95084	4.27539	0.	0.	1.	1.
115.16792	10.73163	83.87040	15.39795	15.89957	0.	0.	2.	0.

APPENDIX 4. SEQUENCE STATISTICS OF SINGLE STRAND RNA

(a) Oligomers Obtained by Degradation of RNA

In Chapter III, we discussed several ways in which oligonucleotides can be prepared by enzymatic or chemical degradation of RNA. In deciding which method to use to prepare a desired oligomer, two factors must to taken into account. First is the theoretical yield of the oligomer. This will depend on the type of degradation chosen, and on the chain length. In addition, it is useful to know how many other oligomers of the same chain length will be present in the reaction mixture. Since oligomers can usually be separated according to base composition, it is helpful to know how many possible compositions there are for each chain length. All of these quantities can be estimated if the RNA chosen for degradation is assumed to have random sequence. The validity of this assumption has been discussed in Chapter IV.

In the case of the base hydrolysis of RNA, it is impossible to predict the yield of oligomers unless the extent of the partial hydrolysis is known. But, if we assume that all bonds are broken at the same rate, we can expect that every possible sequence isomer is present at most stages of the reaction. For oligomers of chain length i, we find that there are 4^i possible sequence isomers, and $(i+3)!/(i)!\,(3)!$ possible compositions.

Consider the total pancreatic RNAase hydrolysis of an RNA which contained X_{py} mole fraction pyrimidines, and X_{pu} mole fraction purines.

For oligomers of chain length i, we would like to know the following two quantities.

X_i is the number of moles of i-mer per total number of moles of oligomers. B_i is the number of moles of monomer residues as i-mer per total number of monomer residues found in the reaction mixture. We find:

There are 2^i possible sequences of chain length i

There are 2i possible compositions with chain length i

$$X_i = X_{py} X_{pu}^{i-1}$$

$$B_i = i X_{py}^2 X_{pu}^{i-1}$$

These equations have been evaluated for chain lengths from one to ten for the degradation products from a hypothetical RNA of infinite chain length which contains equal amounts of A, U, C, and G. These results, shown in Table Ia, are convenient for estimating the yields of oligomers produced by RNAase degradation of RNA. We have also found them useful in determining, from the elution pattern of chromatographic separations, whether the enzyme digestion has gone to completion.

To determine the fragmentation pattern of an RNA digested with T1 RNAase, we must consider X_g, the mole fraction of guanosine, and X_o, the mole fraction of nucleosides other than guanosine. For oligomers of chain length i, we can write, in analogy with the above results, that for T1 RNAase digestion:

There are 3^{i-1} sequences with chain length i

There are (i+1)! / (i-1)! (2)! compositions with chain length i

$$X_i = X_g X_o^{i-1}$$

$$B_i = i X_g^2 X_o^{i-1}$$

These results have been evaluated for chain lengths from one to ten for the conditions described above. The results are tabulated in Table Ib. It is easy to see from the numerical results that T1 digestion of RNA produces many more long chain length oligomers than pancreatic RNAase digestion.

(b) Relation Between Nearest Neighbor Frequencies and Sequence

In Chapter III, we showed that nearest neighbor interactions are sufficient to account for the optical properties of oligonucleotides. This means that, in principle, an analysis of the ORD of an unknown oligomer should permit one to determine the types and quantities of nearest neighbor pairs it contains. Unfortunately, this does not mean that the sequence can be determined by measuring optical properties. For example, we would predict that ApUpApA and ApApUpA should have identical spectra and ORD, yet they have different nucleotide sequences. It is of interest to determine how often a knowledge of the nearest neighbor frequencies will permit a unique reconstruction of the base sequence. Some results of this type have already been discussed in Chapter III. Here we would like to present more details. Of special interest are oligomers obtained from pancreatic RNAase or T1 RNAase

TABLE I

Results of Enzymatic Digestion of Infinite, Random, Equimolar RNA

(a) Pancreatic RNAase

Chain Length	Number of Sequences	X_i	B_i
1	2	1/2	.2500
2	4	1/4	.2500
3	8	1/8	.1875
4	16	1/16	.1250
5	32	1/32	.0781
6	64	1/64	.0469
7	128	1/128	.0273
8	256	1/256	.0156
9	512	1/512	.0088
10	1024	1/1024	.0049
	Fraction of oligomers with chain length of 10 or less		.9941

(b) T1 RNAase

Chain Length	Number of Sequences	X_i	B_i
1	1	.2500	.0625
2	3	.1875	.0938
3	9	.1406	.1055
4	27	.1055	.1055
5	81	.0791	.0989
6	243	.0593	.0890
7	729	.0445	.0779
8	2187	.0334	.0667
9	6561	.0250	.0563
10	19,683	.0188	.0469
	Fraction of oligomers with chain length of 10 or less		.8029

digestion, since these are frequently encountered by workers interested in the determination of nucleic acid sequence.

Unfortunately, we have not been able to find concise algebraic expressions which relate the number of sequence isomers to the number of isomers which are uniquely specified by nearest neighbor frequencies. For oligomers obtained by random fragmentation of RNA, the best approach we have found is to enumerate the various possible kinds of oligomers which cannot be distinguished by nearest neighbor frequencies. These are shown as a function of chain length and composition in Table II. It can easily be seen that the number of indistinguishable sequences rapidly approaches a majority of the possible sequences.

For the simpler case of oligomers obtained by digestion of RNA with specific nucleases, it is easier to find the number of oligomers of a given chain length which have a unique set of nearest neighbor frequencies. It is also possible to determine the number of nearest neighbor categories for a given chain length. This is the number of different possible sets of nearest neighbor frequencies which are consistent with the constraints of base composition and chain length. These results are shown in Table IIIa for oligomers from pancreatic RNAase digestion, and in Table IIIb for oligomers from T1 RNAase digestion. By inspection of the values in Table IIIa, it is noted that the number of nearest neighbor categories for chain length n can be generated by the expression

$$N(n) = N(n-1) + 4(n-2) \quad N \geq 3 .$$

TABLE II

Nearest Neighbor Frequencies for Oligonucleotides

Chain Length	Base Composition†	Number of sequences of a given base composition which do not have unique nearest neighbor frequencies
3	A_3	0
	A_2B	0
	ABC	0
	Total	0
4	A_4	0
	A_3B	24
	A_2B_2	0
	A_2BC	0
	ABCD	0
	Total	24
5	A_5	0
	A_4B	36
	A_3B_2	72
	A_3BC	120
	A_2B_2C	0
	A_2BCD	0
	Total	228
6	A_6	0
	A_5B	48
	A_4B_2	144
	A_4BC	288
	A_3B_3	144
	A_3B_2C	1104
	A_3BCD	192
	$A_2B_2C_2$	0
	A_2B_2CD	0
	Total	1920

†A can be any of the four common bases A, U, C, and G. B is any base different from A; C is any base different from A and B; D is any base different from A, B, and C.

TABLE III

Nearest Neighbor Frequencies of Oligonucleotides

(a) Obtained From Pancreatic RNAase Digests

Chain Length	Number of Sequences	Sequences with unique nearest neighbor frequencies	Number of nearest neighbor categories
1	2	2	2
2	4	4	4
3	8	8	8
4	16	16	16
5	32	24	28
6	64	28	44
7	128	32	64
8	256	36	88

(b) Obtained from T1 RNAase digests

Chain Length	Number of Sequences	Sequences with unique nearest neighbor frequencies	Number of nearest neighbor categories
1	1	1	1
2	3	3	3
3	9	9	9
4	27	27	27
5	81	69	75
6	243	135	186
7	729	207	417
8	2187	?	$\gtrsim$ 909

Similarly, the number of oligomers with unique nearest neighbor frequencies is given by

$$N(n) = N(n-1) + 4 .$$

We have not been able to find similar expressions to generate the statistics of T1 RNAase oligomers.

The results shown in Table III suggest that ORD will be an extremely useful tool for determining the sequence of short oligomers from enzyme digests of RNA. The usefulness of ORD is greater for T1 RNAase oligomers, since the possibility of three nucleosides at each position in the chain leads to many more possible nearest neighbor categories.

APPENDIX 5. OLIGONUCLEOTIDES WHICH CAN INTERACT WITH THEMSELVES TO FORM DOUBLE STRANDED SECTIONS

(a) Intermolecular Interactions - Antiparallel Strands

One way to study the optical properties of double strand RNA is to determine the concentration dependence of the ORD of oligomers which can interact in solution to form double strand section. The simplest possible case occurs when an oligomer is complementary to itself. Thus, if antiparallel strand formation can occur, we might expect GpC to form the double strand $\overrightarrow{GpC}/\overleftarrow{CpG}$, where the arrows indicate the direction of the sugar phosphate chain. If we pick an oligomer at random, what is the probability that it is a perfect self-complement for antiparallel interaction? This can occur only if the oligomer has an even chain length. By inspection, we find that for a chain length i there are $4^{i/2}$ self-complementary oligomers, and 4^i possible sequences. Thus the probability of picking a self-complement is $1/4^{i/2}$. For long oligomers, this is a rare occurrence.

It is hoped that future experiments can determine the minimum number of hydrogen bonded interactions necessary to hold two oligomers together. The weakest condition at all reasonable is that two nearest neighboring hydrogen bond pairs might be sufficient to hold two small oligomers together. Thus, it is of interest to know how many of the possible oligomers of a given chain length can form two near-

est neighboring hydrogen bonds with themselves through intermolecular interaction in the antiparallel mode. This can always occur if the oligomer has two nearest neighbors which are complementary. Define N(i) as the number of oligonucleotides of chain length i which contain at least two adjacent complementary bases. We find

$$N(i) = 4(4^{i-2} + 3/4\ N(i-1)),\ N(1) = 0$$

All oligomers which are perfect self-complements must, of course, contain at least one set of adjacent complementary bases. In addition, it is possible for oligomers which have no nearest neighbor complements to form double strands with two or more adjacent hydrogen bonded pairs. This can occur for chain lengths of five or greater. A typical example is the oligomer ApApCpUpU, which can form four hydrogen bond pairs by an antiparallel interaction. A summary of the above results is shown in Table Ia for chain lengths from one to six.

For short oligomers, it is impossible for an oligomer to be able to form three adjacent hydrogen bond pairs by antiparallel interaction with another oligomer of the same sequence. Thus, the next possibility is to consider oligomers which can form four adjacent base pairs when they interact in an antiparallel fashion. A typical complex formed would be ApC$\overrightarrow{p}$GpU/UpG$\overleftarrow{p}$CpA. The number of oligomers which can form such complexes is shown in Table Ib as a function of chain length. By comparing these results with the ones shown in Table Ia, one can see that it is far more likely for an oligomer to be able to form two adjacent base pairs with itself than four.

TABLE I

Statistics of Intermolecular Interactions in Mixtures Which Contain Only One Oligonucleotide

(a) Antiparallel Double Strands with at Least Two Adjacent Base Pairs

Chain Length	Number of Sequences	Sequences which are exact self-complements	Sequences with adjacent complementary bases, but not exact self-complements	Sequences with no nearest neighbor complements which can form base pairs with a central loop	Total number of sequences which can interact
1	4	0	0	0	0
2	16	4	0	0	4
3	64	0	28	0	28
4	256	16	132	0	148
5	1024	0	700	24	724
6	4096	64	3060	158	3282

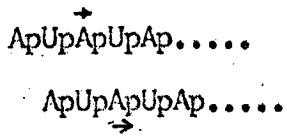
(b) Antiparallel Double Strands with at Least Four Adjacent Base Pairs

1	4	0	0	0	0
2	16	0	0	0	0
3	64	0	0	0	0
4	256	16	0	0	0
5	1024	0	124	0	124
6	4096	64	~660	0	~724

(b) Intermolecular Interactions -

Parallel Strands

While evidence seems to favor antiparallel interaction between oligomers, the possibility of parallel interaction still remains open. Thus, it is of interest to examine the probability that an oligonucleotide chosen at random can interact in solution with another oligomer of the same sequence to form a double strand complex in which both strands are parallel. It is easily shown that no oligomer can be perfectly complementary to itself if interaction occurs in a parallel mode. The closest one can come is the interaction of oligomers of alternating sequence. Thus, ApUpApU... can, in principle, form the parallel double strand complex shown below.



In this complex there will still be two unpaired bases, one at each end. We shall assume, as we did for the case of antiparallel interaction, that at least two adjacent base pairs are necessary to form a double strand parallel complex. It is of interest to determine the number of oligomers of a given chain length which meet the minimum requirement that two parallel base pairs can be formed between two molecules of the same oligomer in solution. This condition is equivalent to asking that the sequence ABA occur at least once in the molecule, where A is complementary to B. We have been able to solve this problem for chain lengths of up to eight, for a given type of

stagger. The stagger is the number of residues one oligomer must be moved relative to the other until two or more adjacent base pairs can be formed. Thus, the example shown on the previous page has a stagger of one. What one would like to know is the probability that an oligomer can form any one of these possible types of staggered interactions. But we have determined only the numbers of ways of forming each single interaction. These results are shown in Table II. The sum of the number of ways of forming each stagger for a given chain length will give some indication of the chance that an oligomer can interact with itself. This is an overestimate, since we have counted those oligomers which can interact in more than one way once for each type of interaction they undergo. By comparing the results shown in Table II and Table Ia, one can see that an oligomer chosen at random is more likely, in principle, to be able to form complexes with antiparallel strands than parallel strands.

TABLE II

Statistics of Intermolecular Interactions in Mixtures Which Contain Only One Oligonucleotide

Parallel Double Strands with at Least Two Adjacent Base Pairs

Chain Length	Stagger	Number of sequences which can interact	Total number of sequences which can interact	Fraction of possible sequences which can interact
2	1	0	0	0
3	1	4		
	2	0		
			4	.063
4	1	28		
	2	16		
			44	.172
5	1	160		
	2	112		
	3	64		
			336	.328
6	1	820		
	2	640		
	3	448		
	4	256		
			2164	.528
7	1	3964		
	2	3280		
	3	2560		
	4	1792		
	5	1024		
			12,620	.770
8	1	18,484		
	2	15,856		
	3	13,120		
	4	10,240		
	5	7,168		
	6	4,096		
			68,964	1.052

APPENDIX 6. PREPARATION OF THE POLYNUCLEOTIDE PHOSPHORYLASE FROM M. LYSODEIKTICUS

The procedure we have used to prepare the polynucleotide phosphorylase from M. lysodeikticus is that published by Steiner and Beers,[A1] for the lysis; Singer and Guss,[A4] for ammonium sulfate and protamine sulfate fractionation; and Singer and O'Brien,[A5] for Sephadex chromatography. The exact conditions we have used are reproduced here as a convenience to anyone who wishes to prepare this enzyme. The reagents used are shown in Table I. The only change we have made in the above procedures is to omit the DEAE-cellulose fractionation recommended by Singer and O'Brien. This step is necessary if enzyme of highest purity is desired, but it causes a loss of about 80% of the total activity of the preparation.

The enzyme is assayed according to the procedure of Singer and O'Brien.[A5] We found that the assay solution is unstable and should be made up just before using. We used $K_2HP^{32}O_4$ of approximately 50,000 cpm per micromole. The assay mixture (total volume 0.1 ml) should contain about 0.01 to 0.04 units of enzyme, 0.1 M tris-HCl, pH 8.2, 5 mM $MgCl_2$, 0.01 M $KHP^{32}O_4$, 1 mM EDTA, and 1 mM poly A. Where necessary, BSA is added to make the total protein concentration above four micrograms per ml. After incubation at 37°C for 15 minutes, the reaction is stopped by adding 1 ml of acid washed norit in 2.5% $HClO_4$. The mixture is stirred occasionally for about ten minutes, then washed onto a millipore filter (HA 0.45 microns, 25 mm diameter). The wet

TABLE I

Reagents Used to Prepare Polynucleotide Phosphorylase

Water - redistilled with a Barnes still.

M. Lysodeikticus cells--acetone dried--Miles Chemical Corp., Lots #36 and #38.

Lysozyme--Worthington Muramidase, Lot #LY 636.

Ammonium Sulfate--Mann #1946, Lot #N2183.

Protamine Sulfate--Nutritional Biochemicals. N.B. Cal. Biochem. protamine sulfate is completely unsatisfactory.

Tris--Eastman 4833, Lot #12A.

2-Mercaptoethanol--Cal. Biochem. 4449, Lot #53428.

$ZnCl_2$--Mallinckrodt 8780.

$MgCl_2$--Mallinckrodt 5958, Lot #JHY.

Poly A--Miles Chemical Corp. Lot #111321.

$H_3P^{32}O_4$--obtained from the Laboratory of Chemical Biodynamics.

Bacterial Alkaline Phosphatase--Worthington Lot #6136.

Bovine Serum Albumin--Pentex, Inc.

UDP, sodium salt--P-L Biochemicals, Lot #1116.

ApC--Cal. Biochem. Lot #45565.

Glycine--Nutritional Biochemicals Corp., Lot #7306.

Phenol Reagent (Folin & Ciocalteu)--Van Waters & Rogers, Lot #10-202.

Blue Dextran 2000--Pharmacia.

Sephadex G-75--Pharmacia.

filter is attached, charcoal side down, to a planchet with four drops of Duco cement, dried under an IR lamp (for one minute) and counted (usually for five minutes) in a thin window gas flow tube with a Baird Atomic Model 123 Scaler.

Protein was determined by a modification of Folin test.[A1] This is a slight variation of the method developed by Lowry et al.,[A3] which had been used in the general references mentioned above.

Reagent A: 2% Na_2CO_3 + 0.02% sodium tartrate, in 0.10 M NaOH

Reagent B: 0.5% $CuSO_4 \cdot 5H_2O$

Reagent C: 50 volumes of A, + 1 volume of B. Make up fresh daily

Reagent E: Phenol reagent (Folin & Ciocalteu), diluted with water until it is 1 N in acid

The test is performed by adding 1 ml of reagent C to 0.2 ml of a protein solution which contains about 100 to 250 micrograms of protein per ml. The mixture is allowed to stand ten minutes, and then 0.1 ml of reagent E is added very rapidly with stirring. The mixture is then allowed to stand for 30 minutes, and the absorbance at 750 mμ is read against an appropriate blank. Standards of Bovine Serum Albumin were used to calibrate the test. These were run every time an unknown was tested. Since we are unsure of the purity of the BSA sample used, all of our protein concentrations have relative accuracy only.

Fraction 1

Measure 200 ml of 0.5% NaCl into a 400 ml beaker. Add 20 gm of dried M. lysodeikticus cells slowly, and stir until an even suspension

is obtained. Bring the pH up to 8.0 to 8.1 by adding 1 N NaOH dropwise. The pH will drift rapidly after this initial titration, and more NaOH usually has to be added until the rate of drift slows to less than .01 pH unit per minute. Bring temperature of the bacterial suspension up to 37°C. Add 50 mg Lysozyme (EC 3.2.1.17) dissolved in 2 ml H_2O, and mix rapidly for a few seconds. Maintain temperature at 37°C. Test the consistency of the suspension occasionally with a glass stirring rod. After about 10 minutes the suspension will set like a custard. Continue stirring for about one minute, until the custard starts to pull away from the walls and looks oily. Immediately add 100 ml of cold saturated ammonium sulfate, and stir to break up the clotted suspension. Let stand 10 minutes, then centrifuge for 30 minutes at 20,000 g. Discard precipitate and measure volume of supernatant (Fraction 1). If this is not between 200 and 225 ml, discard and start again. A small sample of the supernatant can be dialyzed overnight against several changes of distilled water. This is assayed to find the activity of Fraction 1.

Fraction 2

Weigh out 22.3 gm of ammonium sulfate per 100 ml of Fraction 1. Add this over a period of one hour to Fraction 1, either in the cold room or in an ice bath, with continuous slow magnetic stirring. Let stand one hour in the cold room, and centrifuge for 30 minutes at 20,000 g. Pour off supernatant and store, or discard. Dissolve precipitate in 20 ml of 0.1 M tris-HCl, pH 8.1. This takes about one hour and must be performed in the cold room. Dialyze against three

changes (two hours each, or more) of one to two liters of distilled water. The result is Fraction 2.

Fraction 3

Determine the protein concentration of Fraction 2. Dilute it to 10 mg protein per ml, and make final solution 0.1 molar in tris-HCl, pH 8.1 by addint 1.0 M tris pH 8.1, and then distilled water. Add, over a period of one hour, 24.4 gm of ammonium sulfate per 100 ml of diluted Fraction 2. Use magnetic stirrer and ice bath as before. Let stand for one hour in cold room, and centrifuge for 30 minutes at 20,000 g. Collect supernatant and measure volume. Add, over a period of 45 minutes, 8.6 gm ammonium sulfate per 100 ml of supernatant, with stirring, in an ice bath. Let stand for one hour in cold room, and centrifuge at 20,000 g for 30 minutes. Collect precipitate and dissolve in 10 ml of 0.1 M tris-HCl, pH 8.1. Precipitate dissolves easily. Dialyze against three changes (each for two hours or more) of 0.1 M tris-HCl, pH 8.1 containing 0.001 M EDTA. The result is Fraction 3.

Fractions 4 and 5

Measure protein concentration of Fraction 3, and dilute to 4.3 mg of protein per ml with cold distilled water. Cool in an ice bath. Measure 14.4 ml of 0.5% protamine sulfate in water (freshly made every two weeks) per 100 ml of diluted Fraction 3. Add this over a period of one hour to the rapidly stirred diluted Fraction 3. This addition is conveniently performed if the protamine sulfate solution is placed in a large syringe fitted with a small bore teflon needle. Stir for

one hour after the addition, and keep the solution at 0°C. Centrifuge for 10 minutes at 7000 g, and collect the precipitate. Carefully suspend the precipitate in 20 ml of 0.5 M tris-HCl per 100 ml of diluted Fraction 3. Cool in an ice bath and stir for one hour. Centrifuge for 10 minutes at 7000 g, and collect supernatant. Dialyze in batches containing 15 ml or less against 100 volumes of 0.01 M tris-HCl pH 8.1 containing 0.001 M 2-mercaptoethanol. Change dialysis liquid after one hour, and dialyze for one and one half more hours. If any precipitate forms, remove by centrifuging at 7000 g for 10 minutes. The supernatant is Fraction 5.

Fraction 6

Measure volume of Fraction 5 and adjust the solution to pH 6.3 (± 0.1) by adding 1 M acetic acid. If overtitration occurs, adjust by adding 1 N NH_4OH. Keep solution at 0°C. Weigh out 22.6 gm of ammonium sulfate per 100 ml of adjusted Fraction 5. Add this over a period of one hour while keeping the pH of the solution within the above range by adding either acetic acid or ammonium hdyroxide, as needed. Stir for 20 minutes, centrifuge for 10 minutes at 10,000 g, and collect supernatant. Measure volume of supernatant, and cool in an ice bath. Add, with stirring, 12 g. of ammonium sulfate per 100 ml of supernatant over a period of about 45 minutes. Let stand for one hour in cold room, and centrifuge for 10 minutes at 10,000 g. Collect precipitate and dissolve in 1/10 the original volume of Fraction 5 of 0.5 M tris-HCl, pH 8.1. Dialyze against three changes of 200 volumes of 0.01 M tris-HCl, pH 8.1, containing 0.001 M EDTA and 0.001 M 2-

mercaptoethanol. The first two dialyses are for one hour, and the third is for one half hour.

Fraction 7

Measure protein concentration of Fraction 6. Dilute Fraction 6 to 10 mg per ml of protein by adding cold distilled water. At this point it is necessary to do a preliminary test before proceeding. Take four 0.1 ml samples of diluted Fraction 6. Add, respectively, 30, 20, 10, and 5 microliters of 0.1 M $ZnCl_2$ to which a small amount of HCl has been added to dissolve any ZnOCl precipitate.[A8] Let stand for 15 minutes in the cold room, and centrifuge at 10,000 g for 10 minutes. Collect supernatant and measure specific activity. Interpolate among the results to choose the amount of $ZnCl_2$ which should be added to give a high specific activity consistent with recovery of most of the total activity. Frequently, additional preliminary tests will have to be done to narrow in on the correct value. Repeat the above steps using all of Fraction 6 that remains, using the quantity of $ZnCl_2$ per 0.1 ml determined by the preliminary titration. Collect the supernatant as before. Place this on a 39 x 3.8 cm G-75 Sephadex column which has been washed with 400 ml of 0.01 M tris-HCl containing 0.001 M EDTA, 0.001 M 2-mercaptoethanol, and 0.25 M NaCl. Elute with the same buffer; protein starts to appear at the void volume (determined with blue dextran). Most of the activity will appear in the first 18 ml of the protein peak. The enzyme in any stage from Fraction 1 to Fraction 7 can be stored frozen. Typical activities and concentrations of the various fractions are shown in Table II.

TABLE II

Typical Results of Presparation of Polynucleotide Phosphorylase

Fraction	Specific Activity (units/mg protein)	Units/ml	Protein mg/ml	Total Activity* (units)	A_{280}/A_{260}
1	0.22	0.45	2	99	0.84
2	0.23	4.81	20.8	192	0.89
3	0.37**	6.38**	17.1**	89**	1.02**
5	1.00**	7.58**	6.9**	93**	1.53**
6	1.30	22.1	17.0	87	1.67
7	20.6†	12.4†	0.6†	37†§	

*Based on 20 g of M. Lysodeikticus cells as starting material.

**Average of two runs.

†Fraction at peak of protein elution.

§Total of 55 units were recovered.

Synthesis of Oligonucleotides

Thus far we have only very preliminary results. In an attempt to prepare the trimer ApCpU we followed the procedure of Leder, Singer and Brimacombe explicitly,[A2] but the products of the reaction included only long polymer, and unreacted monomer and dimer. When we attempted the same reaction, using the procedure of Thach and Doty,[A7] we obtained as products four additional substances which should correspond to ApCpU, ApCpUpU, ApCpUpUpU, and ApCpUpUpUpU. None of these has been fully characterized yet, but the UV spectra and ORD of the fraction which should be ApCpU correspond very well with reported[A2] and calculated values. We are at a loss to explain why, in our hands, one procedure works while the other does not; but, at this juncture, it is fair to say that we have insufficiently tested both methods.

REFERENCES

A1. Glitz, D., Quantitative Chemical Techniques of Histochemistry and Cytochemistry, Vol. II, Interscience, New York, 1963, p. 153.

A2. Leder, P., Singer, M. F. and Brimacombe, R. L., Biochemistry 4, 1561 (1965).

A3. Lowry, O. H.,Rosebrough, N. J., Farr, A. L. and Randall, R. J., J. Biol. Chem. 193, 265 (1961).

A4. Singer, M. F. and Cuss, J. K., J. Biol. Chem. 237, 182 (1962).

A5. Singer, M. F. and O'Brien, B. M., J. Biol. Chem. 238, 328 (1963).

A6. Steiner, R. F. and Beers, Jr., R. F., Polynucleotides, Elsevier Press, Amsterdam, 1961.

A7. Thach, R. E. and Doty, P., Science 148, 632 (1965).

A8. Williamson, S., private communication.

REFERENCES

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

2. Apgar, J., Everett, G. A., and Holley, R. W., Proc. Natl. Acad. Sci. U.S., 53, 546 (1965).

3. Applequist, J., and Damle, V., J. Am. Chem. Soc. 87, 1450 (1965).

4. Armstrong, A., Hagopian, H., Ingram, V. M., Sjöquist, I., and Sjöquist, J., Biochemistry 3, 1194 (1964).

5. Bak, T., Contributions to the Theory of Chemical Kinetics, W. A. Benjamin, New York, 1963, p. 21.

6. Bautz, E.K.F., and Bautz, E. A., Proc. Natl. Acad. Sci. U.S., 52, 1476 (1964).

7. Bautz, E.K.F., and Hall, B. D., Proc. Natl. Acad. Sci. U.S., 48, 400 (1962).

8. Bautz, E.K.F., and Heding, L., Biochemistry 3, 1010 (1964).

9. Beer, M., and Moudrianakis, E. N., Proc. Natl. Acad. Sci. U.S., 48, 409 (1962).

10. Berg, P., Lagerkvist, V., and Dieckmann, M., J. Mol. Biol. 5, 159 (1962).

11. Bergquist, P. C., and Scott, J. F., Biochim. Biophys. Acta 87, 199 (1964).

12. Bishop, D.H.L., and Bradley, D. E., Biochem. J. 95, 82 (1965).

13. Bjork, W., J. Biol. Chem. 238, 2487 (1963).

14. Bock, R., private communication.

15. Bockstahler, L. E., and Kaesberg, P., J. Mol. Biol. 13, 127 (1965).

16. Boedtker, H., Biochim. Biophys. Acta 32, 519 (1959).

17. Boedtker, H., J. Mol. Biol. 2, 171 (1960).

18. Brahms, J., J. Mol. Biol. 11, 785 (1965).

19. Brahms, J., and Mommaerts, W.F.H.M., J. Mol. Biol. 10, 73 (1964).

20. Brimacombe, R., Trupin, J., Nirenberg, M., Leder, P., Bernfield, M., and Jaouni, T., Proc. Natl. Acad. Sci. U.S., 54, 954 (1965).

21. Brode, W. R., Chemical Spectroscopy, John Wiley, New York, 1939, p. 195.

22. Brown, G. L., Prog. Nucleic Acid Res. 2, 260 (1963).

23. Burney, A., and Marbaix, G., Biochim. Biophys. Acta 103, 409 (1965).

24. Burton, K., Essays in Biochemistry 1, 57 (1965).

25. Bush, C. A., Ph.D. thesis, University of California, 1965.

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

27. Cantor, C. R., Tinoco, Jr., I., and Peller, L., Biopolymers 2, 51 (1964).

28. Chamberlin, M., private communication.

29. Chamberlin, M., Baldwin, R. L., and Berg, P., J. Mol. Biol. 7, 334 (1963).

30. Chamberlin, M. J., and Patterson, D. L., J. Mol. Biol. 12, 410 (1965).

31. Chan, S. I., Schweizer, M. P., Ts'o, P.O.P., and Helmkamp, G. K., J. Am. Chem. Soc. 86, 4182 (1964).

32. Clark, Jr., J. M., Eyzaguirre, J. P., and Gunther, J. K., J. Biol. Chem. 239, 1877 (1964).

33. Cohn, W. E., and Uziel, M., Information Exchange Group #7, Memo #59.

34. Crestfield, A. M., Smith, K. C., and Allen, F. W., J. Biol. Chem. 216, 185 (1955).

35. Crick, F.H.C., Information Exchange Group #7, Memo #14.

36. Crick, F.H.C., Barnett, L., Brenner, S., and Watts-Tobin, R. J., Nature 192, 1227 (1961).

37. Davis, R., private communication.

38. Davis, S. L., Ph.D. thesis, University of California, 1965.

39. Dekker, C., private communication.

40. DeVoe, H., and Tinoco, Jr., I., J. Mol. Biol. 4, 500 (1962).

41. DeVoe, H., and Tinoco, Jr., I., J. Mol. Biol. 4, 518 (1962).

42. Dimroth, K., and Witzel, H., Liebigs. Ann. Chem. 620, 109 (1954).

43. Doctor, B. P., Connelly, C. M., Rushizky, G. W., and Sober, H. A., J. Biol. Chem. 238, 3985 (1963).

44. Doctor, B. P., and McCormick, G. J., Biochemistry 4, 49 (1965).

45. Donohue, J., and Trueblood, K. N., J. Mol. Biol. 2, 363 (1960).

46. Drobnik, J., and Kleinwächter, V., Biochem. Biophys. Res. Commun. 21, 366 (1965).

47. Duesberg, P. H., and Robinson, W. S., Proc. Natl. Acad. Sci. U.S., 54, 794 (1965).

48. Dwight, H. B., Tables of Integrals, 4th ed., John Wiley, New York, 1961, p. 38.

49. Edstrom, J. E., Biochim. Biophys. Acta 80, 399 (1964).

50. Egami, F., Progr. Theoret. Phys. (Kyoto), 17, 92 (1961).

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

52. Elson, D., and Chargaff, E., Biochim. Biophys. Acta 17, 369 (1955).

53. Englander, S. W., and Englander, J. J., Proc. Natl. Acad. Sci. U.S., 53, 370 (1965).

54. Fasman, G. D., Lindblow, C., and Grossman, L., Biochemistry 3, 1015 (1964).

55. Fasman, G. D., Lindblow, C., and Seaman, E., J. Mol. Biol. 12, 630 (1965).

56. Felsenfeld, G., and Cantoni, G. L., Proc. Natl. Acad. Sci. U.S., 51, 818 (1964).

57. Felsenfeld, G., and Hirschman, S. Z., J. Mol. Biol. 13, 407 (1965).

58. Fiers, W., DeWachter, R., Lepoutre, L., and Vandendriessche, L., J. Mol. Biol. 13, 451 (1965).

59. Fiers, W. E., and Khorana, H. G., J. Biol. Chem. 238, 2780 (1963).

60. Fiers, W. E., and Khorana, H. G., J. Biol. Chem. 238, 2789 (1963).

61. Fiers, W., Lepoutre, L., and Vandendriessche, L., J. Mol. Biol. 13, 432 (1965).

62. Fiers, W., and Sinsheimer, R. L., J. Mol. Biol. 5, 408 (1962).

63. Flory, P. J., J. Am. Chem. Soc. 62, 1561 (1940).

64. Fraenkel-Conrat, H., Scientific American, 211, No. 4, 46 (1964).

65. Freifelder, D., Kleinschmidt, A. K., and Sinsheimer, R. L., Science 146, 254 (1964).

66. Fresco, J. R., Informational Macromolecules, ed. Vogel, H., Academic Press, New York, 1963.

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

68. Fresco, J. R., Klotz, L. C., and Richards, E. G., Cold Spring Harbor Symposia, 28, 83 (1963).

69. Fruton, J. S., and Simmonds, S., General Biochemistry, John Wiley, New York, 1953, p. 251.

70. Gellert, M., Lipsett, M. N., and Davies, D. R., Proc. Natl. Acad. Sci. U.S., 48, 2013 (1962).

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

72. Gilham, P. T., J. Am. Chem. Soc. 86, 4982 (1964).

73. Gilham, P. T., and Robinson, W. E., J. Am. Chem. Soc. 86, 4985 (1964).

74. Glaubiger, D. L., Ph.D. thesis, University of California, 1965.

75. Glitz, D. G., and Dekker, C. A., Biochemistry 2, 1185 (1963).

76. Gomatos, P. J., Krug, R. M., and Tamm, I., J. Mol. Biol. 9, 193 (1964).

77. Gomatos, P. J., Krug, R. M., and Tamm, I., J. Mol. Biol. 13, 802 (1965).

78. Gomatos, P. J., and Tamm, I., Proc. Natl. Acad. Sci. U.S., 50, 878 (1963).

79. Grunberg-Manago, M., Prog. Nucleic Acid Res. 1, 93 (1963).

80. Guschlbauer, W., Richards, E. G., Beurling, K., Adams, A., and Fresco, J. R., Biochemistry 4, 964 (1965).

81. Hadjivassiliou, A., and Brawerman, G., Biochim. Biophys. Acta 103, 211 (1965).

82. Haggis, G. H., Michie, D., Muir, A. R., Roberts, K. B., and Walker, P.M.B., Introduction to Molecular Biology, John Wiley, New York, 1964.

83. Handbook of Chemistry and Physics, 40th ed., 1958, p. 286.

84. Harris, R., Prog. Nucleic Acid Res. 2, 30 (1963).

85. Hartmann, Jr., K. A. and Rich, A., J. Am. Chem. Soc. 87, 2033 (1965).

86. Haschemeyer, A.E.V., and Rich, A., private communication.

87. Haschemeyer, A.E.V., and Sobell, H. M., Proc. Natl. Acad. Sci. U.S., 50, 872 (1963).

88. Haschemeyer, A.E.V., and Sobell, H. M., Nature 202, 969 (1964).

89. Haschemeyer, R., Singer, B., and Fraenkel-Conrat, H., Proc. Natl. Acad. Sci. U.S., 45, 313 (1959).

90. Haselkorn, R., and Fox, C. F., J. Mol. Biol. 13, 780 (1965).

91. Hearst, J., private communication.

92. Henley, D., Lindahl, T., and Fresco, J., Information Exchange Group #7, Memo #112.

93. Hill, T. L., Statistical Thermodynamics, Addison-Wesley, Reading, Mass., 1960, p. 238.

94. Hilmoe, R. J., J. Biol. Chem. 235, 2117 (1960).

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

96. Holley, R. W., Science 150, 921 (1965).

97. 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).

98. Holley, R. W., Everett, G. A., Madison, J. T., and Zamir, A., J. Biol. Chem. 240, 2122 (1965).

99. Holley, R. W., Madison, J. T., and Zamir, A., Biochem. Biophys. Res. Commun. 17, 389 (1965).

100. Howard, F. B., Frazier, J., Lipsett, M. N., and Miles, H. T., Biochem. Biophys. Res. Commun. 17, 93 (1964).

101. Huang, R. C., and Bonner, J., Proc. Natl. Acad. Sci. U.S., 54, 960 (1965).

102. Hurwitz, J., and August, J. T., Prog. Nucleic Acid Res. 1, 75 (1963).

103. Ingram, R. B., and Baldwin, R. L., J. Mol. Biol. 5, 172 (1962).

104. Ingram, V. M., and Pierce, J. G., Biochemistry 1, 581 (1962).

105. Jaskunas, S. R., private communication.

106. Jordon, D. O., The Chemistry of Nucleic Acids, Butterworths, London, 1960, p. 91.

107. Jukes, T., private communication.

108. Keir, H. M., Mathog, R. H., and Carter, C. E., Biochemistry 3, 1188 (1964).

109. Keller, E. B., Biochem. Biophys. Res. Commun. 17, 412 (1964).

110. Kerr, I. M., Pratt, E. A., and Lehman, I. R., Biochem. Biophys. Res. Commun. 20, 154 (1965).

111. Koerner, J. F., and Sinsheimer, R. L., J. Biol. Chem. 228, 1049 (1957).

112. Krieg, D. R., Prog. Nucleic Acid Res. 2, 125 (1963).

113. Krug, R. M., Gomatos, P. J., and Tamm, I., J. Mol. Biol. 12, 872 (1965).

114. Lagerkvist, V., and Berg, P., J. Mol. Biol. 5, 139 (1962).

115. Lamborg, M. R., and Zamecnik, P. C., Biochem. Biophys. Res. Commun. 20, 328 (1965).

116. Lamborg, M. R., Zamecnik, P. C., Li, T., Kägi, J., and Vallee, B. L., Biochemistry 4, 63 (1965).

117. Lane, B. G., Biochemistry 4, 212 (1965).

118. Langridge, R., and Gomatos, P. J., Science 141, 694 (1963).

119. Langridge, R., and Marmur, J., Science 143, 1450 (1964).

120. Lapidot, Y., and Khorana, H. G., J. Am. Chem. Soc. 85, 3852 (1963).

121. Leder, P., Singer, M. F., and Brimacombe, R.L.C., Biochemistry 4, 1561 (1965).

122. Lee, J. C., and Gilham, P. T., J. Am. Chem. Soc. 87, 4000 (1965).

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

124. Lee, S., McMullen, D., Brown, G. L., and Stokes, A. R., Biochem. J. 94, 314 (1965).

125. Lehman, I. R., Prog. Nucleic Acid Res. 2, 83 (1963).

126. Lehman, I. R., and Nussbaum, A. L., J. Biol. Chem. 239, 2628 (1964).

127. Lin, C. Y., Urry, D. W., and Eyring, H., Biochem. Biophys. Res. Commun. 17, 642 (1964).

128. Lindahl, T., Henley, D. D., and Fresco, J. R., J. Am. Chem. Soc. 87, 4961 (1965).

129. Lipsett, M. N., Proc. Natl. Acad. Sci. U.S., 46, 445 (1960).

130. Lipsett, M. N., J. Biol. Chem. 239, 1250 (1964).

131. Lipsett, M. N., J. Biol. Chem. 239, 1256 (1964).

132. Lipsett, M. N., and Heppel, L. A., J. Am. Chem. Soc. 85, 118 (1963).

133. Lipsett, M. N., Heppel, L. A., and Bradley, D. F., Biochim. Biophys. Acta 41, 175 (1960).

134. Lipsett, M. N., Heppel, L. A., and Bradley, D. F., J. Biol. Chem. 236, 857 (1961).

135. Lloyd, D., private communication.

136. Lohrmann, R., and Khorana, H. G., J. Am. Chem. Soc. 86, 4188 (1965).

137. Luzzati, V., Mathis, A., Masson, F., and Witz, J., J. Mol. Biol. 10, 28 (1964).

138. McMullen, D., private communication.

139. Magee, Jr., W. S., Gibbs, J. H., and Newell, G. F., J. Chem. Phys. 43, 2115 (1965).

140. Magee, Jr., W. S., Gibbs, J. H., and Zimm, B. H., Biopolymers 1, 133 (1963).

141. Mahler, H. R., Kline, B., and Mehrotra, B. D., J. Mol. Biol. 9, 801 (1964).

142. Mahler, H., Moore, W., and Thompson, R., Information Exchange Group #7, Memo #87.

143. Mandeles, S., private communication.

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

145. Mangiantini, M. T., Tecce, G., Toschi, G., and Trentalance, A., Biochim. Biophys. Acta 103, 252 (1965).

146. Markham, R., The Viruses, 2, 33 (1959).

147. Massoulie, J., and Michelson, A. M., C. R. Acad. Sci. Paris, 259, 2923 (1964).

148. Matus, A. I., Ralph, R. K., and Mandel, H. G., J. Mol. Biol. 10, 295 (1965).

149. Mellema, J., Moudrianakis, E. N., and Beer, M., presented before the 9th Annual Meeting of the Biophysical Society, 1965.

150. Meuron-Landolt, M., and Privat de Garihle, M., Biochim. Biophys. Acta 91, 433 (1964).

151. Michelson, A. M., The Chemistry of Nucleosides and Nucleotides, Academic Press, New York, 1963.

152. Miles, H. T., Proc. Natl. Acad. Sci. U.S., 51, 1104 (1964).

153. Miles, H. T., and Frazier, J., Biochem. Biophys. Res. Commun. 14, 21 (1964).

154. Miller, W. G., and Alberty, R. A., J. Am. Chem. Soc. 80, 5146 (1958).

155. Miura, K., J. Biochem. (Tokyo), 55, 452 (1964).

156. Moudrianakis, E. N., and Beer, M., Biochim. Biophys. Acta 95, 23 (1965).

157. Moudrianakis, E. N., and Beer, M., Proc. Natl. Acad. Sci. U.S., 53, 564 (1965).

158. Nash, H. A., and Bradley, D. F., Biopolymers 3, 261 (1965).

159. Neu, H. C., and Heppel, L. A., J. Biol. Chem. 239, 2927 (1964).

160. Newmark, R. A., Ph.D. thesis, University of California, 1964.

161. Nihei, T., and Cantoni, G. L., Biochim. Biophys. Acta 61, 463 (1962).

162. Nihei, T., and Cantoni, G. L., J. Biol. Chem. 238, 3991 (1963).

163. Nirenberg, M., and Leder, P., Science 145, 1399 (1964).

164. Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F., and O'Neal, C., Proc. Natl. Acad. Sci. U.S., 53, 1161 (1965).

165. Nozo, K., Harada, Y., Ynasa, S., and Honja, I., Biochim. Biophys. Acta 103, 517 (1965).

166. Pabst Laboratories, Circular OR-10 (1956).

167. Peller, L., and Alberty, R. A., J. Am. Chem. Soc. 81, 5907 (1959).

168. Penswick, J. R., and Holley, R. W., Proc. Natl. Acad. Sci. U.S., 53, 543 (1965).

169. Perrin, D. D., Austral. J. Chem. 16, 572 (1963).

170. Pratt, A. W., Toal, J. N., Rushizky, G. W., and Sober, H. A., Biochemistry 3, 1831 (1964).

171. Printz, M. P., and von Hipple, P. H., Proc, Natl. Acad. Sci. U.S., 53, 363 (1965).

172. Privat de Garihle, M., Les Nucleases, Hermann, Paris, 1964, p. 30.

173. Privat de Garihle, M., Cunningham, L., Laurila, U. R., and Laskowski, M., J. Biol. Chem. 224, 751 (1957).

174. Pochon, F., and Michelson, A. M., Proc. Natl. Acad. Sci. U.S., 53, 1425 (1965).

175. Ralph, R. K., Conners, W. J., and Khorana, H. G., J. Am. Chem. Soc. 84, 2265 (1962).

176. Rauth, A. M., Biophys. J. 5, 257 (1965).

177. Razzell, W. E., and Khorana, H. G., J. Biol. Chem. 234, 2105 (1959).

178. Razzell, W. E., and Khorana, H. G., J. Biol. Chem. 234, 2114 (1959).

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

180. Rice, W. E., and Bock, R. M., J. Theor. Biol. 4, 260 (1963).

181. Rich, A., Davies, D. R., Crick, F.H.C., and Watson, J. D., J. Mol. Biol. 3, 71 (1961).

182. Richards, E. G., Flessel, C. P., and Fresco, J. R., Biopolymers 1, 431 (1963).

183. Robinson, W. S., Pitkamen, A., and Rubin, H., Proc. Natl. Acad. Sci. U.S., 54, 137 (1965).

184. Ross, P. D., and Scruggs, R. L., Biopolymers 3, 491 (1965).

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

186. Rushizky, G. W., and Knight, C. A., Proc. Natl. Acad. Sci. U.S., 46, 945 (1960).

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

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

189. Rushizky, G. W., and Sober, H. A., J. Biol. Chem. 237, 2883 (1962).

190. Rushizky, G. W., and Sober, H. A., Biochem. Biophys. Res. Commun. 14, 276 (1964).

191. Rushizky, G. W., Sober, H. A., Connelly, C. M., and Doctor, B. P., Biochem. Biophys. Res. Commun. 18, 489 (1965).

192. Samejima, T., and Yang, J. T., J. Biol. Chem. 240, 2094 (1965).

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

194. Sarin, P. S., and Zamecnik, P. C., Biochem. Biophys. Res. Commun. 20, 400 (1965).

195. Sarkar, P. K., and Yang, J. T., J. Biol. Chem. 240, 2088 (1965).

196. Sarkar, P. K., and Yang, J. T., Biochemistry 4, 1238 (1965).

197. Sarkar, P. K., and Yang, J. T., Biochem. Biophys. Res. Commun. 20, 346 (1965).

198. Sauer, K., private communication.

199. Schildkraut, C., and Lifson, S., Biopolymers 3, 195 (1965).

200. Schweizer, M. P., Chan, S. I., and Ts'o, P.O.P., J. Am. Chem. Soc. 87, 5241 (1965).

201. Sedat, J. W., and Hall, J. B., J. Mol. Biol. 12, 174 (1965).

202. Shapiro, M. B., Merril, C. R., Bradley, D. F., and Mosimann, J. E., Science 150, 918 (1965).

203. Shefter, E., Barlow, M., Sparks, R., and Trueblood, K., J. Am. Chem. Soc. 86, 1872 (1964).

204. Sinanoglu, O., and Abdulnur, S., J. Photochem. Photobiol. 3, 333 (1964).

205. Singer, B., and Fraenkel-Conrat, H., Biochim. Biophys. Acta 72, 534 (1963).

206. Singer, B., and Fraenkel-Conrat, H., Biochim. Biophys. Acta 76, 143 (1963).

207. Singer, M. F., and Guss, J. K., J. Biol. Chem. 237, 182 (1962).

208. Singer, M. F., and O'Brien, B. M., J. Biol. Chem. 238, 328 (1963).

209. Singer, M. F., and Tolbert, G., Science 145, 593 (1964).

210. Singer, M. F., and Tolbert, G., Biochemistry 4, 1319 (1965).

211. Singh, H., and Lane, B. G., Can. J. Biochem. 42, 1011 (1964).

212. Sinha, N. K., Enger, M. D., and Kaesberg, P., J. Mol. Biol. 12, 299 (1965).

213. Sinha, N. K., Fujimura, R. K., and Kaesberg, P., J. Mol. Biol. 11, 64 (1965).

214. Söll, D., and Khorana, H. G., J. Am. Chem. Soc. 87, 350 (1965).

215. Spatz, H. C., and Baldwin, R. L., J. Mol. Biol. 11, 213 (1965).

216. Spencer, M., and Poole, F., J. Mol. Biol. 11, 314 (1965).

217. Spirin, A. S., Prog. Nucleic Acid Res. 1, 301 (1963).

218. Staehelin, M., Biochim. Biophys. Acta 49, 11 (1961).

219. Staehelin, M., Biochim. Biophys. Acta 49, 20 (1961).

220. Staehelin, M., Biochim. Biophys. Acta 49, 27 (1961).

221. Staehelin, M., Prog. Nucleic Acid Res. 2, 170 (1963).

222. Stanley, Jr., W., Ph.D. thesis, University of Wisconsin, 1964.

223. Stanley, Jr., W. M., and Bock, R. M., Biochemistry 4, 1302 (1965).

224. Steiner, R. F., and Beers, Jr., R. F., Polynucleotides, Elsevier, Amsterdam, 1961.

225. Steinschneider, A., and Fraenkel-Conrat, H., Science 150, 386 (1965).

226. Strauss, J. H., and Sinsheimer, R. L., J. Mol. Biol. 7, 43 (1963).

227. Stretton, A.O.W., and Brenner, S., J. Mol. Biol. 12, 451 (1965).

228. Sugiyama, T., and Fraenkel-Conrat, H., Proc. Natl. Acad. Sci. U.S., 47, 1393 (1961).

229. Takanami, M., private communication.

230. Tanford, C., Physical Chemistry of Macromolecules, John Wiley, New York, 1961.

231. Technicon Instruments Corp., Bulletin No. RNA-1.

232. Thach, R. E., and Doty, P., Science 147, 1311 (1965).

233. Thach, R. E., and Doty, P., Science 148, 632 (1965).

234. Thach, R. E., and Sundararajan, T. A., Proc. Natl. Acad. Sci. U.S., 53, 1021 (1965).

235. Tinoco, Jr., I., J. Am. Chem. Soc. 82, 4785 (1960).

236. Tinoco, Jr., I., J. Am. Chem. Soc. 86, 297 (1964).

237. Tinoco, Jr., I., Molecular Biophysics, ed. B. Pullman and M. Weissbluth, Academic Press, New York, 1965, p. 269.

238. Tinoco, Jr., I., private communication.

239. Tomita, K. I., and Rich, A., Nature 201, 1160 (1964).

240. Tomlinson, B., private communication.

241. Tomlinson, R. V., and Tener, G. M., Biochemistry 2, 697 (1963).

242. Ts'o, P.O.P., and Chan, S. I., J. Am. Chem. Soc. 86, 4176 (1964).

243. Tsugita, A., and Fraenkel-Conrat, H., J. Mol. Biol. 4, 73 (1962).

244. Tuppy, H., and Küchler, E., Biochim. Biophys. Acta 80, 669 (1964).

245. Ulbricht, T.L.V., private communication.

246. Van Holde, K. E., Brahms, J., and Michelson, A. M., J. Mol. Biol. 12, 726 (1965).

247. Vournakis, J. N., and Scheraga, H. A., private communication.

248. Vournakis, J. N., Scheraga, H. A., Rushizky, G. W., and Sober, H. A., Biopolymers, in press.

249. Wacker, A., and Lodemann, E., Angew. Chem. Int. 4, 150 (1965).

250. Wahl, P., private communication.

251. Warshaw, M. M., private communication.

252. Warshaw, M. M., Ph.D. thesis, University of California, 1965.

253. Warshaw, M. M., Bush, C. A., and Tinoco, Jr., I., Biochem. Biophys. Res. Commun. 18, 633 (1965).

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

255. Watson, J. D., Molecular Biology of the Gene, W. A. Benjamin, New York, 1965.

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

257. Whitfeld, P. R., J. Biol. Chem. 237, 2865 (1962).

258. Williams, E. J., Sung, S., and Laskowski, Sr., M., J. Biol. Chem. 236, 1130 (1961).

259. Witz, J., Hirth, L., and Luzzati, V., J. Mol. Biol. 11, 613 (1965).

260. Witz, J., and Luzzati, V., J. Mol. Biol. 11, 620 (1965).

261. Witzel, H., and Barnard, E. A., Biochem. Biophys. Res. Commun. 7, 295 (1962).

262. Worthington Enzyme Catalogue.

263. Ycas, M., Nature 188, 209 (1960).

264. Yolles, R., Ph.D. thesis, University of California, 1964.

265. Yu, C. T., and Zamecnik, P. C., Biochim. Biophys. Acta 45, 148 (1960).

266. Yudkin, M. D., and Davis, B., J. Mol. Biol. 12, 193 (1965).

267. Zachau, H. G., unpublished results.