Lawrence Berkeley National Laboratory
Recent Work

Title
CIRCULAR DICHROISM STUDIES ON SYNTHETIC DOUBLE HELICAL OLIGORIBONUCLEOTIDES

Permalink
https://escholarship.org/uc/item/8qs8x8qc

Author
Borer, Philip Noel.

Publication Date
1972-10-01

eScholarship.org Powered by the California Digital Library
University of California

LBL-1365
c. 1

CIRCULAR DICHROISM STUDIES ON SYNTHETIC DOUBLE HELICAL OLIGORIBONUCLEOTIDES

Philip Noel Borer

(Ph. D. Thesis)

October 1972

RECEIVED
LAWRENCE
RADIATION LABORATORY

DEC 18 1972

LIBRARY AND
DOCUMENTS SECTION

AEC Contract No. W-7405-eng-48

For Reference

Not to be taken from this room



LBL-1365
c. 1

DISCLAIMER

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

0 0 0 0 3 9 0 0 0 4 8

This thesis is dedicated to Jane, the best wife and friend a man could have, and to Scotty, the happiest, most lovable little boy in my world.

ABSTRACT

The circular dichroism (CD) of a set of complementary oligoribonucleotides forming Watson-Crick base-paired double helices is studied. Also, the circular dichroism and temperature dependence of hypochromicity of a set of ribo-oligomers, $A_6C_mU_6$, forming unimolecular hairpin loops is reported. The techniques for synthesis of ribo-oligomers of defined sequence are described, with particular emphasis on the use of primer dependent polynucleotide phosphorylase.

The set of double strand CD spectra are analyzed in terms of the first-neighbor (nearest-neighbor) only approximation. The set of double strands has sufficient variety to analyze for each of the 10 unique first-neighbor contributions to the CD of double strand RNA. A matrix technique for extracting the contributions from the measured spectra is discussed. First neighbor calculations of the CD of high-polymeric synthetic RNA show good agreement with measured spectra.

Analysis of the temperature dependence of hypochromicity for the hairpin loops yields information on the stability of such regions in natural RNA. Loops of six to seven cytosine residues are the most stable. Stability drops sharply as the loop size decreases beyond five. The circular dichroism results show that the C residues have approximately the same average conformation as in

0 0 0 0 3 9 0 0 5 4 9

single strand oligo C_n for loops larger than five. Smaller loops appear to have an altered conformation.

Acknowledgements.

I especially thank Olke Uhlenbeck for making experimental work as exciting as it is for me today, and Nacho Tinoco for his guidance, helpful discussions, and giving me enough free rein to make some mistakes but gain confidence in myself as a researcher.

A few of the older heads that have bumped mine in the right direction deserve particular thanks, Dave Shirley, Jim Wang, John Hearst, and Ken Sauer.

There are two people who have worked long and hard with me to assemble the data in this thesis, Barbara Dengler and David Koh. Also, without Barbara's cheerful personality and David's great Chinese cooking, life around the lab would have been much the poorer. More recently, Cemal Kemal, Mei Chou, and George Miljanich have participated in some of the experiments. Their contributions are much appreciated.

I gratefully acknowledge the guidance and friendship of the postdoctoral fellows who have taken the time to learn and share their knowledge with me, Olke Uhlenbeck, Don Gray, Fritz Allen, Kashy Javaherian, and Werner Hug.

Some of the people to whom I owe a special kinship in the struggle, "In partial fulfillment of the requirements..." are Dave Lloyd, Carl Formoso, Arlene Blum, Mark Levine, Eric Wickstrom, Carol Cech, Lily Sun, and

0 0 0 0 3 9 0 0 6 5 0

Burt Dorman. Burt's help in using his Calcomp plotting program is greatly appreciated. Without it, I would still be plotting data or would have long since given up.

The booby prizes for general lack of helpfulness go to the PDP 8/s computer for being slow and frequently out of order, and the systems programmers responsible for the operation of the CDC 7600 and 6600 computers. Their sole purpose in life seems to be revising a perfectly good operating system and in the process crashing the computer for a day or two at a critical point in my analysis. Also, the LBL paper tape reader deserves a Bronx cheer for mutilating data tapes or simply not reading them.

My wife, Jane, and Tanya Kmit who helped in the preparation of this manuscript must be given proper credit for the fine job they did.

The author was supported for most of his tenure as a National Institutes of Health predoctoral fellow under USPH grant GM 41393. He was also availed of the facilities of the Chemical Biodynamics Laboratory of the Lawrence Berkeley Laboratory which is supported by the AEC.

Title: CIRCULAR DICHROISM STUDIES ON SYNTHETIC DOUBLE HELICAL OLIGORIBONUCLEOTIDES

Page

I. Introduction 1

A. Why Should Nucleic Acid Structure Concern Us? 1

B. How Can We Probe Nucleic Acid Structure? . . . 2

1. Primary Structure Can Be Known for Reasonably Short Nucleic Acids 2

2. Base-pairing Schemes Can Be Used to Guess Conformation 3

3. Chemical Attack Yields Some Structural Information 4

4. Physical Methods Can Evaluate Possible Conformations and Suggest Others 4

a. Gross Characteristics Can Be Obtained By Studying Natural Polymers 5

b. Polymer Properties Can Be Understood by Studying Suitable Model Compounds . 5

i. The First Neighbor Approximation and "Suitable Model Compounds . . 6

ii. Limitations on Quantitative Results 13

c. Useful Physical Methods 20

i. Hydrodynamic Properties 20

ii. Spectroscopy 21

iii. Other Methods 34

0 0 0 0 3 9 0 0 3 5 1

Page

II. Materials and Methods 40

A. General Considerations on Synthesis of Block Oligoribonucleotides 40

B. Synthetic Reactions 41

1. Partial Hydrolysis of Homopolymers 41

2. Specific Nuclease Action on Mixed Copolymers 47

3. Polynucleotide Phosphorylase Reactions . . 48

a. Preparation of High Molecular Weight Polymers 49

b. Primer Dependent Kinase Reactions -- Simple Block Additions 52

c. Nuclease Assisted Kinase Reaction . . 55

d. Nuclease Assisted Double Kinase Reaction 55

e. Kinase/phosphorylase Equilibrium Reaction 57

f. Gilham Reaction -- 3'-blocked Diphosphates 61

g. Carbodiimide Reaction -- Blocking Nuclease Action 62

4. Other Nuclease Reactions 63

C. Separation Techniques 64

D. Identification Techniques 79

E. Treatment of Hypochromicity Data 82

F. Extinction Coefficient Determinations and Error Estimates 86

G. Circular Dichroism and Ultraviolet Absorption Measurements 91

Appendix

II - A Sample Time Trial and Synthesis for A_4U_m 100

II - B Primer Dependent PNPase Assay 108

Page

III. Results on Double Helices 115

A. The Series $\overrightarrow{A_nU_n}$ Has Only $(\overrightarrow{AA})_d$ and $(\overrightarrow{AU})_d$ Couples 115

B. $\overrightarrow{A_4UAU_4}$ Is Slightly More Complex 118

C. $\overrightarrow{A_nCU_m}$ + $\overrightarrow{A_mGU_n}$ Have One GC Base Pair 128

D. $\overrightarrow{A_nGCU_n}$ Have Two G-C Base Pairs 128

E. $\overrightarrow{A_nCGU_n}$ Have Two G-C Base Pairs But Different First Neighbor Couples 129

F. $\overrightarrow{U_2CGA_2}$ Is Also A Sequence Isomer Of $\overrightarrow{A_2GCU_2}$ . . 134

G. $\overrightarrow{A_3CUG}$ + $\overrightarrow{CAGU_3}$ Have Two Non-adjacent G-C Base Pairs 137

H. $\overrightarrow{A_4G_2}$ + $\overrightarrow{C_2U_4}$ Have a G-C Base Pair On The End . 137

I. $\overrightarrow{A_4CG}$ + $\overrightarrow{CGU_4}$ Have a $(\overrightarrow{CG})_d$ Couple On The End . . 141

IV. Discussion of Results 144

A. Analysis For Individual First Neighbor Contributions 144

1. Theory 144

2. Double Strand CD Contributions 148

a. Linear Combinations of Spectra Giving the Double Strand Contributions . . . 154

b. Reconstituting the Input Spectra From The $[\overrightarrow{JK}]_d$ 156

c. Predicting Other Double Strand Spectra 157

3. Single Strand CD Contributions 170

B. Suggestions for Further Work 171

0 0 0 0 3 9 0 0 8 5 2

Page

V. The Stability of RNA Hairpin Loops: $A_6C_mU_6$ . . 174

A. Introduction 175

B. Materials and Methods 176
1. Synthesis 176
2. Measurements 177
3. Treatment of Melting Data 180

C. Thermodynamic Results 185
1. Single Strand to Complete Loop 185
2. Loop Initiation 187

D. Thermodynamic Discussion 196

E. Circular Dichroism 199

F. Conclusion 211

Appendix 1:
Computer Programs 215
A. Program LISTIN 215
B. Program SUPSPC 218
C. Program SMOOSP 219
D. Program GETNSM 225
E. Program CDCONT 236
F. Program INTERP 252
G. Program INTERP2 257
H. Program DIFFSPC 261

Appendix 2:
Double Strand First Neighbor Contributions . . . 266

0 0 0 0 3 9 0 0 8 5 3

CHAPTER I

Introduction

A. Why Should Nucleic Acid Structure Concern Us?

Nucleic acids are the master molecules in all known forms of life. That is, their sequence and conformation carry all of the information necessary to sustain the stockpile of genetic information an organism has accumulated over the eons. In addition, they direct the complex process of synthesizing the proteins through which all life processes are accomplished. From the standpoint of pure curiosity and awe it behooves us to understand these molecules, but there are also interesting engineering possibilities. Perhaps we can correct malfunctioning systems, make functioning ones more efficient, inhibit or kill deleterious organisms, or even, with fear and trepidation, design new systems for our own uses. These goals can be reached most easily with a thorough understanding of the molecular processes involved. The functions of these processes are indivisibly united with nucleic acid structure.

According to present knowledge about molecular biology, the linear sequence of bases in DNA determines the unique character of an organism. In principle, aside from environmental factors which are not well understood,

knowledge of the information contained in a DNA would allow us to make an a priori prediction of its development. So far, this knowledge has been very difficult to obtain so molecular biologists have had to content themselves with studying simpler problems.

One of these problems is the resolution of the way in which RNA carries out its many functions. Whatever RNA-related process we study we find a definite relation to the three-dimensional structure of the RNA. It is the purpose of the experiments described in this dissertation to provide the molecular biologist with some refined tools to explore RNA conformation. James Watson's book[1] gives a lucid description of the present state of knowledge about the relation of nucleic acid structure and function.

B. How Can We Probe Nucleic Acid Structure?

1. Primary Structure Can Be Known for Reasonably Short Nucleic Acids

Sanger and his co-workers[2] have developed techniques for determining the linear sequence of bases in RNA by analysis of the fragments obtained from enzymatic digests. Since the procedures are adaptable to the use of ^{32}P labelled RNA's, one is able to work with the very small amounts of material one gets from limited enzymatic digests of the RNA to be sequenced. However, the work

0 0 0 0 5 9 0 0 8 5 4

gets very tedious and the separations difficult if the subject RNA gets longer than about 100 bases. Recently it has become possible to sequence longer mRNA's which can be replicated.[3] The technique involves starting the replication on all of the RNA's in the sample at a specific location, adding a pulse of ^{32}P labelled triphosphates, and then resuming synthesis on all the RNA's in synchrony. The labelled sections of RNA are then analyzed by conventional Sanger techniques. The lack of total synchrony eventually defeats this procedure for very long RNA's.

2. Base-pairing Schemes Can Be Used to Guess Conformation

Once the primary structure has been determined it is possible to make educated guesses about the secondary structure. On the lowest level of these guesses, one simply tries to maximize the number of base pairs. Further refinements on this procedure have recently been published which give more realistic weighting to G - C vs. A - U base pairs and account for the destabilizing influence of various types of loops.[4,5] It is also apparent that the sequence of bases in both base paired regions and single strand loops influences stability and should be introduced in appropriate parameters to evaluate possible conformations.

3. Chemical Attack Yields Some Structural Information

Cramer[6] in a review on tRNA structure lists several chemical reactions quite specific for certain bases when they are in single strands. One carries out these reactions under conditions where the RNA is in a "native" conformation and then determines which bases are covalently modified. If a base suspected to be in a single strand region does not react the proposed structure is suspect, but it may also be that the tertiary folding of the RNA protects that single strand region from chemical attack.

4. Physical Methods Can Evaluate Possible Conformations and Suggest Others

The use of physical methods to learn about conformational aspects of biopolymers is attractive because only small amounts of material are generally needed and it is treated non-destructively. We will consider the various kinds of studies and the information they yield in a later section. But, in general, one makes reasonable models to explain the physical property manifested by the molecule and tests them by studying other properties or varying environmental parameters that affect the conformation. Thus, possible structures are evaluated. Often an empirical result will suggest new conformational possibilities if the property is sufficiently well under-

0 0 0 0 3 9 0 0 0 5 5

stood in application to other systems.

a. Gross Characteristics Can Be Obtained By Studying Natural Polymers

Certain conformational aspects of biomolecules are determinable by studying the ones that are actually synthesized in intact organisms. There are many physical properties, for instance, that are proportional to molecular weight and overall shape and density of biopolymers. Thus, hydrodynamic studies on natural polymers have yielded useful scales for obtaining the molecular weight of unknown materials. In addition to the hydrodynamic properties we should list various chromatographic phenomena as contributing to our knowledge of gross conformational knowledge.

b. Polymer Properties Can Be Understood by Studying Suitable Model Compounds

As one moves from a more macroscopic examination of polymer properties down to finer structural points, more sophisticated models are necessary. The compounds needed for more detailed explanation of conformation are usually not found in nature. Natural compounds generally have too many of the interactions that are of interest to serve as simple enough models. Therefore, molecular biologists usually resort to studying the most simple of natural polymers, studying their partial degradation products, or synthesizing suitably simple yet pertinent

compounds. It is this latter tack we have used in this study.

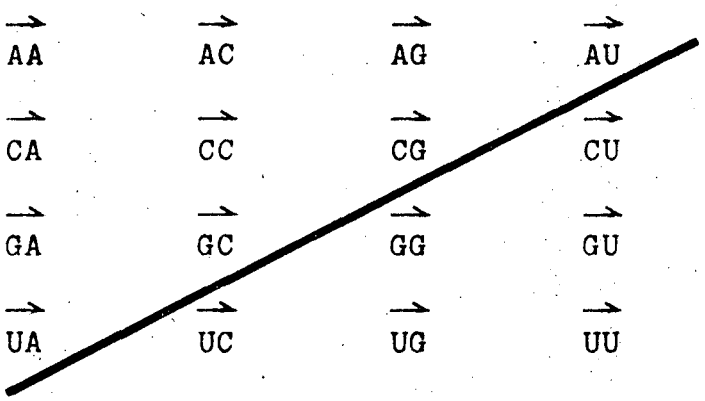
1. The First Neighbor Approximation and "Suitable Model Compounds"

We now explore a particular application of model studies in the properties of ribonucleic acids dependent on base sequence. The particular emphasis will be on properties influenced by the electronic transitions in the bases and stability effects. Gray and Tinoco[7] have extensively described the various levels of approximation used in explaining the sequence dependent properties of polynucleotides and their applicability to specific phenomena. Briefly, the so-called zeroth-neighbor approximation holds that the property depends only on base composition. The first-neighbor (or nearest-neighbor) approximation is adequate to describe physical quantities that depend only on the dinucleotide frequency in the polymer. These definitions are extended to the second-neighbor (next-nearest-neighbor), third neighbor, and higher levels when the phenomenon has a linear dependence on the trinucleotide, tetranucleotide frequency, etc. For instance, bouyant densities are quite nearly proportional to the base composition[8] making this a zeroth neighbor property. The ultraviolet absorption and circular dichroism of most RNA and DNA single strands[9-14] and double strands[14,15] can be adequately described by a first-neighbor approximation. It has been shown that in

0 0 0 0 3 9 0 0 8 5 6

the case of the ribose trimers AUG, AUA, and GUG the first-neighbor approximation does not give close agreement with the experimental CD[14], indicating that the U residue might loop out allowing the purines to form a "hydrophobic" stack and the more soluble U better contact with the water. In cases with internal single strand U residues, it seems necessary to invoke a second neighbor approximation to explain their behavior.

We now develop some further details of the first-neighbor approach since the main body of the experiments in this thesis supports the notion that nucleic acid optical properties and stability depends primarily on first-neighbor frequencies. For the four common bases in RNA, the following 16 single strand first-neighbor combinations (referred to hereafter as single strand couples) can occur:



where the arrow represents the 5' - 3' direction.

For double strands, the number of couples is reduced to ten if only Watson-Crick type base pairs are allowed. In other words, the double strand couple $\begin{matrix}\overrightarrow{CA}\\ \underleftarrow{GU}\end{matrix}$ is the same as $\begin{matrix}\overrightarrow{UG}\\ \underleftarrow{AC}\end{matrix}$. We use a condensed notation for the double strand couples where we specify only one strand and the complement is implied. Thus, for double strands, $\overrightarrow{AG} = \begin{matrix}\overrightarrow{AG}\\ \underleftarrow{UC}\end{matrix}$. The couples beneath the heavy line above are redundant and are not used.

The mathematics of using the first-neighbor approximation are quite simple but it is easy to make mistakes if a systematic approach is not followed. This is particularly true in the case of double strands. Therefore, we present a slightly different approach from that used in the past.[7,9] The particular advantages of this "interaction" approach are the ease with which it can be applied, the correspondence between the first-neighbor "contributions" and the polymer case, and its direct intuitive appeal.

We first develop the equations with simple single strands of oligo A_n, and double strand oligo A_n:oligo U_n. If a molecule of A_n exhibits the physical property, $P(A_n)$, the nearest-neighbor approximation says that

(1-1) $$P(A_n) = nP_A + (n-1)I_s(\overrightarrow{AA}) .$$

In other words, the property is due to the sum of monomer properties plus a sum of the interactions between adjacent single strand bases of the form $I_s(\overrightarrow{AA})$. We now define a nearest neighbor "contribution", $[\overrightarrow{AA}]_s$

0 0 0 0 3 9 0 0 0 5 7

(1-2) $$[\overrightarrow{AA}]_s = I_s(\overrightarrow{AA}) + P_A .$$

$[\overrightarrow{AA}]_s$ is the contribution of a dimer stack in the middle of an oligomer -- the sum of the interaction and one-half of the monomer property from each of the first-neighboring bases. Substituting (1-1) into (1-2)

(1-3) $$[\overrightarrow{AA}]_s = \frac{P(A_n) - P_A}{n-1} .$$

For double strand $(A_n:U_n)$ the situation is parallel, except the "monomers" are of the type $\binom{A}{U}$ and the "dimers", $\binom{\overrightarrow{AA}}{\underleftarrow{UU}}$.

(1-4) $$P(A_n:U_n) = nP\binom{A}{U} + (n-1)I_d\binom{\overrightarrow{AA}}{\underleftarrow{UU}} .$$

and

(1-5) $$[\overrightarrow{AA}]_d = I_d\binom{\overrightarrow{AA}}{\underleftarrow{UU}} + P\binom{A}{U} .$$

Then

(1-6) $$[\overrightarrow{AA}]_d = \frac{P(A_n:U_n) - P\binom{A}{U}}{n-1}$$

These equations are correct for one molecule or a mole of molecules. Frequently in work with nucleic acids we normalize the physical properties to a value per mole of nucleotide residue to facilitate comparisons between monomers, dimers, trimers, tec., and polymers. Let us indicate this molar residue property by $\overline{P}$. Then $\overline{P} = \frac{P}{m}$ where m is the total number of bases contributing to P. Substituting in eq. (1-3), m = n and $P(A_n) = n\overline{P}(A_n)$. $(P_A = \overline{P}_A)$.

(1-7) $$[\overrightarrow{AA}]_s = \frac{n\overline{P}(A_n) - P_A}{n-1} .$$

For eq. (1-6), $m = 2n$, $P(A_n:U_n) = 2n\bar{P}(A_n:U_n)$, and

$$P\begin{pmatrix}A\\U\end{pmatrix} = 2\bar{P}\begin{pmatrix}A\\U\end{pmatrix},$$

$$(1\text{-}8) \qquad [\overrightarrow{AA}]_d = \frac{2n\bar{P}(A_n:U_n) - 2\bar{P}\begin{pmatrix}A\\U\end{pmatrix}}{n-1}.$$

We usually set $\bar{P}\begin{pmatrix}A\\U\end{pmatrix} = 1/2(P_A + P_U)$. This is usually not bad since the ends make only a small contribution to an oligomer property. Also, for CD we will see later that nearly coplanar bases contribute very little to the CD. Then eq. (1-8) becomes

$$(1\text{-}9) \qquad [\overrightarrow{AA}]_d = \frac{2n\bar{P}(A_n:U_n) - (P_A + P_U)}{n-1}.$$

Two interesting cases of the application of equations (1-7) and (1-9) are for the dimer ($n = 2$) and polymer ($n \to \infty$).

Single strand:

$n = 2$ (1-10) $[\overrightarrow{AA}]_s = 2\bar{P}(\overline{ApA}) - P_A$

$n \to \infty$ (1-11) $[\overrightarrow{AA}]_s = \bar{P}(\text{poly A})$

Double strand:

$n = 2$ (1-12) $[\overrightarrow{AA}]_d = 4\bar{P}\left(\begin{matrix}\overline{ApA}\\ \underline{UpU}\end{matrix}\right) - (P_A + P_U)$

$n \to \infty$ (1-13) $[\overrightarrow{AA}]_d = 2\bar{P}(\text{poly A:poly U})$

There is a nice correlation between the contributions and polymer properties. Blum[16] made the correlation even better by defining her double strand contributions to be half of the definition in eq. (1-13). This makes certain calculations slightly more cumbersome and the proper coefficients somewhat less obvious, however. Note that

the coefficients on the right side of the single strand equations sum to one, each $[\overrightarrow{AA}]_s$ is a contribution from one mole of residues. Likewise, $[\overrightarrow{AA}]_d$ is a contribution from two moles. This gives us a sum rule that can be used as a check on more complex calculations -- the sum of single strand coefficients and twice the coefficients of the double strand contributions must be equal for both sides of an equation. This rule will be illustrated below. First we generalize the forms for arbitrary sequences of length n.

$$(1\text{-}14) \qquad n\bar{P}(\text{oligo}_n) = \sum_{J=1}^{4} \sum_{K=1}^{4} Q^{s}_{\overrightarrow{JK}} [\overrightarrow{JK}]_s + \frac{1}{2} \sum_{1}^{2} (P_1)$$

$$(1\text{-}15) \qquad 2n\bar{P}(\text{oligo}_n : \text{oligo}_n^{*}) = \sum_{J=1}^{4} \sum_{K=1}^{5-J} Q^{d}_{\overrightarrow{JK}} [\overrightarrow{JK}]_d + \frac{1}{2} \sum_{1}^{4} (P_1),$$

where the sums are over the four bases A,C,G, and U in that order; $Q^{s}_{\overrightarrow{JK}}$ or $Q^{d}_{\overrightarrow{JK}}$ is the number of the dimer sequence $\overrightarrow{JK}$ (counted along one strand only in the double strand case); the P_1 are for the end bases; and oligo_n^{*} is the complement of oligo_n. The following examples illustrate the use of equations (1-14) and (1-15).

$$5\bar{P}(\overrightarrow{GAUUU}) = [\overrightarrow{GA}]_s + [\overrightarrow{AU}]_s + 2[\overrightarrow{UU}]_s + 1/2(P_G + P_U)$$

$$5 = 1 + 1 + 2 + 1/2 + 1/2$$

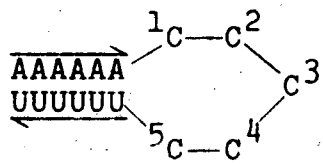
(sum rule obeyed)

$$10\overline{P}\ \begin{matrix}\overrightarrow{GAUUU}\\ \underleftarrow{CUAAA}\end{matrix} =$$

$$[\overrightarrow{GA}]_d + [\overrightarrow{AU}]_d + 2[\overrightarrow{AA}]_d + 1/2(P_A + P_C + P_G + P_U)$$

$$10 = 2 \times 1 + 2 \times 1 + 2 \times 2 + 1/2 + 1/2 + 1/2 + 1/2$$

(sum rule obeyed)



$$17\ \begin{matrix}\overrightarrow{AAAAAA}\\ \underleftarrow{UUUUUU}\end{matrix} =$$

$$5[\overrightarrow{AA}]_d + 1/2(P_A + P_U) + 4[\overrightarrow{CC}]_s + [\overrightarrow{AC}]_s + [\overrightarrow{CU}]_s$$

$$17 = 5 \times 2 + 1/2 + 1/2 + 4 + 1 + 1$$

(sum rule obeyed)

In the last example, an approximation to the interface between loop and double strand was made. It assumes that the AC and CU single strand interactions are nearly the same as in the base stacks at the interface and that there is no interaction between C^1 and the nearest U, nor between C^5 and the nearest A.

We have given some hints as to what we consider suitable model compounds but we now try to present them coherently. On the surface it appears that dimers, oligomers or polymers would be equally good models. We argue in favor of an oligomer approach since in the dimers, none of the bases is really in the center of a stack such as occurs in natural RNA's. It seems likely that bases in the center of a stack will be less influenced by solvent effects. Also, in the double strand case the "dimer" is not stable under anything approaching physiological conditions. The polymers may not be suitable for calcu-

0 0 0 3 5 9 0 0 5 3 8

lations with smaller molecules because long range effects may influence their geometry. It has been demonstrated for DNA's, at least, that some of the simpler model polymers do apparently have geometry other than B-form.[15] Also, there is not as versatile a synthetic method available for polymers of complex sequence as for oligomers. Oligomers which are linear combinations of the various $[\overrightarrow{JK}]_d$ can be quite easily made and they are quite stable at near physiological conditions. In some fortunate instances, one can even do better with oligomers than a nearest neighbor calculation, and actually account for some longer range interactions. For instance the physical properties of hairpin loops, $A_nC_mU_n$ can be approximated by an appropriate sum of double strand (oligo A_n:oligo U_n), oligo C_m, ApC, CpU, and the monomers. The two long oligomers account for some of the longer range interactions in the hairpin loop.

ii. Limitations on Quantitative Results

The basic limitation on calculations performed with the first neighbor approximation is that first-neighbor interactions do not tell the whole story. Calculations dealing with ordinary ultraviolet absorption are usually quite good, since from theoretically sound arguments, the oscillator strength and therefore the absorbance should depend on r^{-3} where r is the distance between the bases. The ultraviolet CD and ORD of nucleic acids de-

pend in theory on r^{-2} and therefore should be less well explained by a first-neighbor analysis. Cantor[13] obtained excellent qualitative agreement between measured and calculated trinucleoside diphosphate ORD spectra. There is also good qualitative correlation in the shapes of homopolymer ORD[10] and CD[14]. The CD of many double strand DNA's is also quite well characterized by a first-neighbor analysis.[15] Some of the trimer and polymer spectra are reproduced in figures 1-1 and 1-2. One can clearly see that even with a perfectly accurate and noise-free spectrometer, however, that first-neighbor calculations will never give greatly detailed sequence information for very long oligomers. Not only is there the distance problem, it is apparent that a variety of conformations do exist that depend on sequence, viz., the strange appearance of trimer CD spectra which have central uridine residues alluded to above, and the strange CD of DNA's with all pyrimidines on one strand and all purines on the other.[15] In addition, other factors might cause DNA to take on the A-form or C-form geometry, such as packing in a phage or chromosome. Finally, of course, there are limitations on the accuracy to which a measurement can be made. In Chapter II, we estimate that the maximum absolute error in a single CD measurement at the long wavelength maximum to be about ±5% and the maximum relative error (all measurements made on the same machine and concentrations all determined in the same way) to be

Figure 1-1.

The ORD of trinucleoside diphosphates at pH 7: measured (————) and first neighbor calculated (--------). Taken from Cantor et al.[13]

Figure 1-2.

The CD of two single strand ribopolymers at pH 7: measured (————) and first neighbor calculated (--------). Spectra taken by D.M. Gray (unpublished).

a. poly ribo A

b. poly ribo U

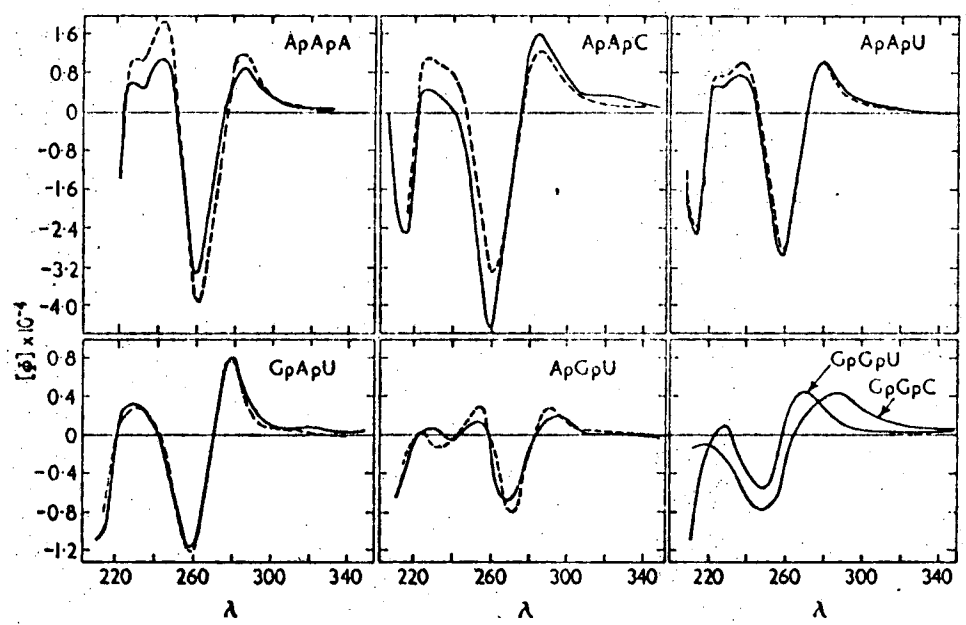

ApApA
ApApC
ApApU
GpApU
ApGpU
GpGpU
GpGpC
[φ] x10⁻⁴
λ
220
260
300
340

Figure 1-1

0 0 0 0 3 9 0 0 8 5 9

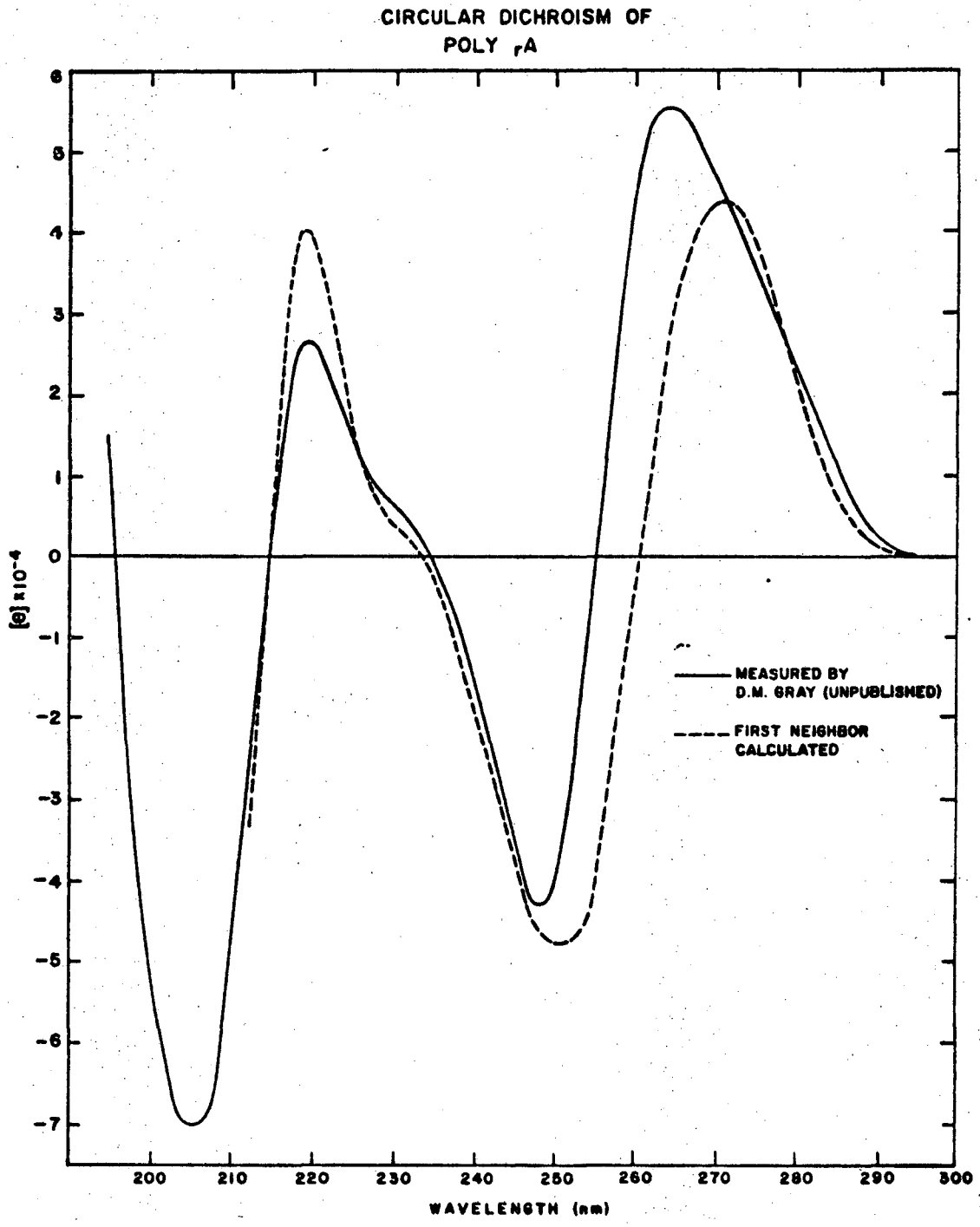


Figure 1-2a

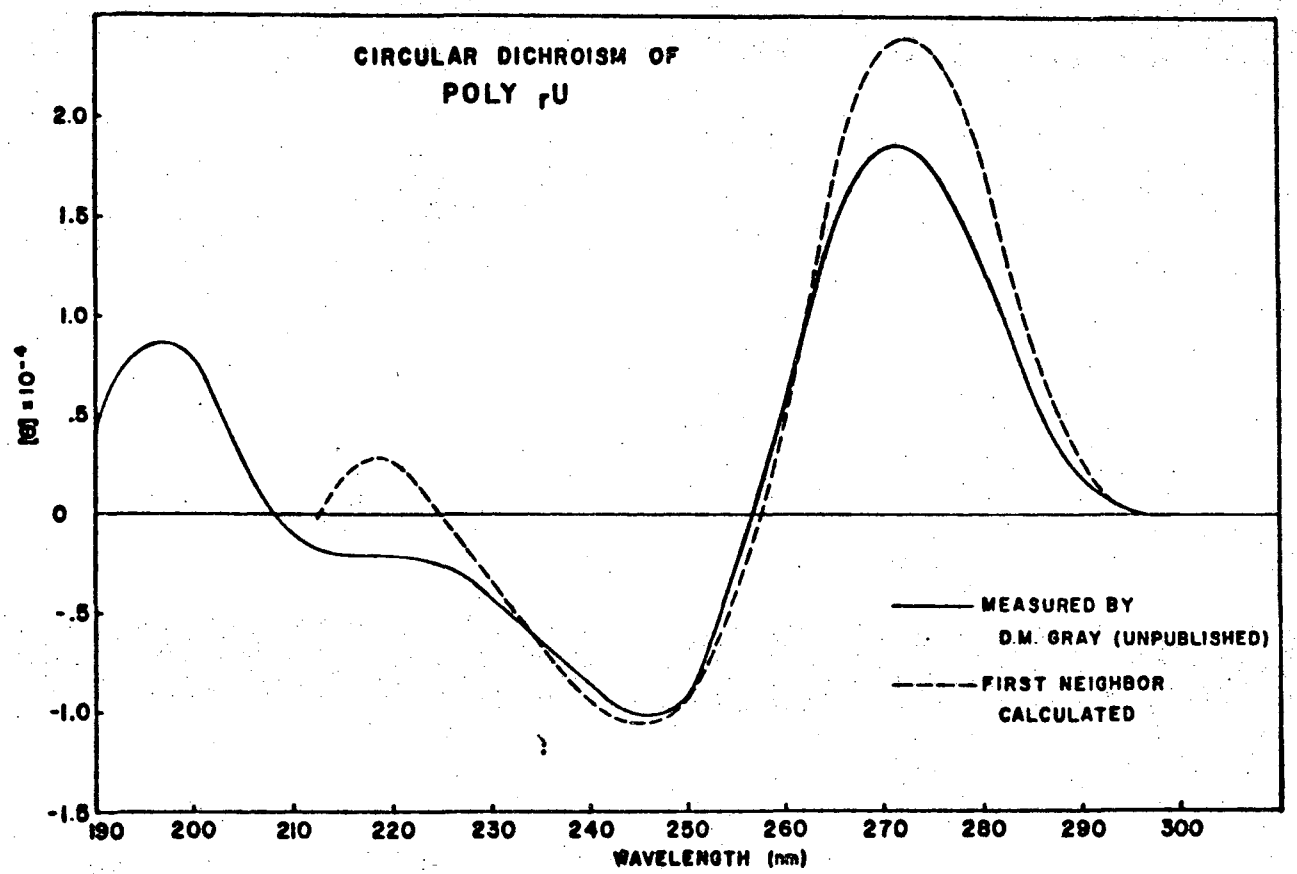

CIRCULAR DICHROISM OF
POLY rU
2.0
1.5
1.0
.5
0
-.5
-1.0
-1.5
190
200
210
220
230
240
250
260
270
280
290
300
WAVELENGTH (nm)
MEASURED BY
D.M. GRAY (UNPUBLISHED)
FIRST NEIGHBOR
CALCULATED

Figure 1-2b

0 0 0 0 3 9 0 0 8 6 0

± 3%. Gray, et al.[15] took the standard deviation of five scans on two different preparations of M. luteus DNA and obtained a standard deviation at the long wavelength maximum characteristic of an error of ± 5%. Their standard deviations were about the same throughout the spectral region below the maximum but above 210 nm where it was about ± 10%. The standard deviation decreases above 280 nm. It should be noted that oligo-RNA solutions have about five times the CD of DNA samples at the same molar residue concentration. Therefore, the contributions to the error due to instrumental causes will be smaller for the RNA's. Of course, concentration errors will still be of the same magnitude. Experimental errors are less serious for the single strand case where the $[\overrightarrow{JK}]_s$ terms can be obtained from a simple equation, eg., eq. (1-10) with relatively few measured quantities. If eq. (1-13) is used for the double strands, again we need deal with only a few measured quantities.[15] However, we present evidence (see Chapters III and IV) that the polymers probably do not well represent the conformation of short double helical RNA. Extracting the $[\overrightarrow{JK}]_d$ terms (or using linear combinations thereof) from double helical oligomers is more prone to error since the oligomers one can make are rather restricted in sequence. This means that the contributions from some first-neighbor combinations are quite dilute compared with others, and are affected more by experimental errors. This is one of the arguments

that compelled us to synthesize more than the minimum number of oligomers to isolate each contribution.

c. Useful Physical Methods

The greatest single appeal to the molecular biologist of the use of physical procedures is that they are usually non-destructive. If he has spent six months preparing or isolating a tiny amount of a compound, he usually is very willing to shoot a beam of light through his sample as long as it is not destroyed, even though he doesn't really understand the method (or results!). On the other side of the coin, biological macromolecules are particularly difficult to study because of their complexity making physical properties one of the few probes available for finding out about their structure. The criteria, then, for a particularly "useful physical method" are that it be easy to understand and apply, applicable to small amounts of material, and sensitive to absolute conformation or small changes in conformation. The following methods discussed more or less satisfy these requirements.

1. Hydrodynamic Properties

The study of the transport properties of biological molecules, for example viscosity and ultracentrifugation, was one of the first methods to yield information on macromolecular organization. Transport phenomena are

0 0 0 0 3 9 0 0 8 6 |

still extensively used to isolate and study biopolymers. The type of information obtained is usually more on the level of gross characteristics, such as molecular weight, bouyant densities, and state of inter- and intramolecular aggregation.

ii. Spectroscopy

This rather small heading covers the considerable ground of visible, ultraviolet, and infrared absorption; circular dichroism; optical rotatory dispersion; nuclear magnetic resonance; and scattering phenomena. With the exception of the latter, which gives information about gross physical properties, these are probes of localized groups within a biopolymer. They may reflect gross structural changes, but do so only because the environment of many of the groups has changed.

Infrared[17,18] and Raman[19] studies have been made on nucleic acids which give information about the number and kind of base pairs present. Such conclusions are possible because the O-H and N-H stretching force constants are lowered by hydrogen bonding. A principal disadvantage of the method is that the extinction coefficients for IR absorption are low, requiring large amounts of material for measurements. Also the exact interpretation of results is difficult because of the present lack of enough studies on suitable models; e.g., poly A: poly U and poly (alt AU): poly (alt AU) have unexplained

spectral differences.

Nuclear magnetic resonance is a very powerful tool for conformational analysis in chemistry. It is being applied more and more extensively to the study of biological molecules. Again, perhaps its biggest disadvantage is the large amount of material required. Several papers have been published on the pmr of tRNA in water,[20,21] bases pairing in DMSO,[22] and nucleosides pairing in non-aqueous systems.[23] The evidence is that the protons involved in base pairing are shifted far downfield from the others, into the region from -10 to -15 ppm. Other sutdies on base stacking with bases and nucleosides[24-27] and dinucleoside phosphates[28-32] show an upfield shift.

Another application of nmr spectroscopy to nucleic acids may be the use of ^{13}C - nmr.[33-39] Its insertion as a specific probe in a specific location of an RNA would be very useful for 3-dimensional structure determinations. Synthesizing oligo-RNA's with one ^{13}C labelled base would be extremely easy with the oligomer synthesizing system described in this dissertation and a source of ^{13}C enriched nucleoside diphosphates.

So far the techniques we have discussed are those which have not been used in this study but show considerable promise in further elucidating the structure of double helical oligonucleotides and natural RNA. We now examine the kind of information we can gain from ultra-

0 0 0 0 3 9 0 0 8 6 2

violet absorption and circular dichroism. We will not discuss optical rotatory dispersion explicitly since it contains no different information from circular dichroism.[40]

The magnitude of absorption for electronic transitions in nucleic acids is primarily determined by the transitions in each of the bases separately. That this is true is easily verified by experiment; a polymer has approximately the same magnitude and shape of absorption as an appropriate mixture of its constituent monomers. However, a polynucleotide does absorb less than its monomers in the normal ultraviolet region from 200 to 320 nm, and this effect is called hypochromism (at a particular wavelength the effect is called hypochromicity). On the other hand, the magnitude and shape of circular dichroism bands for aggregates of bases is usually much different from the CD of the individual monomers. A thorough explanation of all of the details of nucleic acid spectra is impossible at the present time due to the present state of knowledge about these compounds, due both to experimental difficulties in getting data on the short wavelength transitions and the inaccuracy of the approximations needed to calculate the wavefunctions for such large systems. However, there are several qualitative statements that can be made to rationalize the measured spectra, especially for simple models. First we shall cover some of the background theory of

Figure 1-3.

The absorption spectra of GpA and its constituent monomers. Taken from Warshaw.[41]

0 0 0 0 3 9 0 0 3 6 3

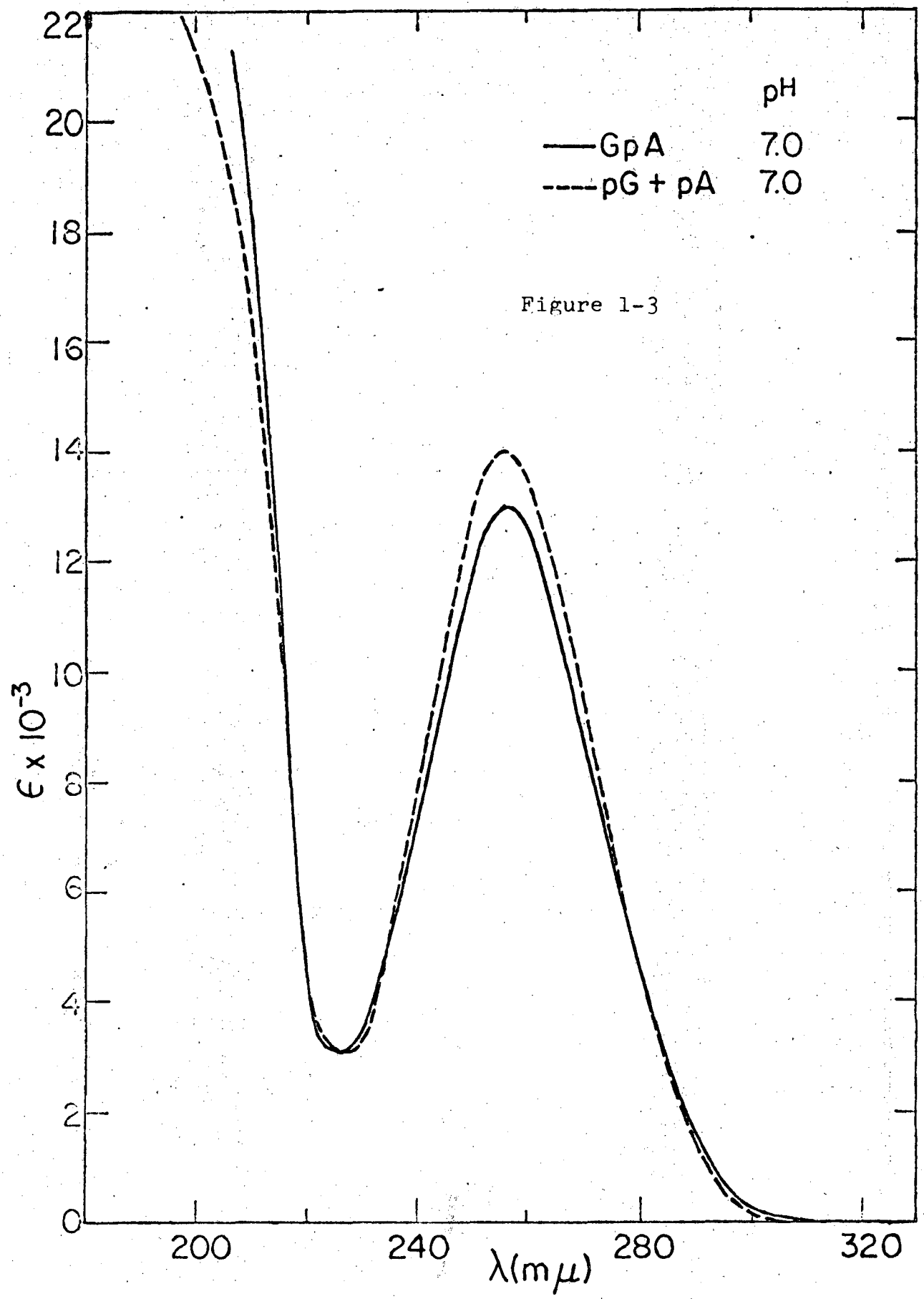


Figure 1-3

optical properties.

Since the monomer ground and excited states are about equally shifted upon formation of an aggregate dimer or other polymer there are usually small spectral frequency shifts between a monomer and polymer. This is illustrated in figure 1-3 for a typical dimer and its constituent monomers.[41] A good approximation for nucleic acids is that there is no exchange of electrons between the bases and no overlap in the wavefunctions. Then for a polymer of N monomer units, the polymer wavefunctions, ψ_i, for the excited states take the form

$$\psi_i = \sum_{k=1}^{N} C_{ik}\phi_k, \quad i,k = 1,N \tag{1-16}$$

where the ϕ_k are monomer wavefunctions and the C_{ik} are constants. The total number of excited states is conserved, but the excitation is not localized on a particular monomer, hence the term exciton states. Some of the excited states have higher or lower energy than the monomer but the average value is the same.

Hypochromism and circular dichroism are caused by the coulombic interactions between separate groups when an electronic transition occurs. These coulombic terms take several forms including dipole - induced dipole, dipole - dipole, and dipole - charge interactions. When a transition occurs, the electron density changes. This change is usually approximated as a transition electric dipole moment, $\vec{\mu}_{oA}$,

0 0 0 0 3 9 0 0 3 6 4

(1-17) $\vec{\mu}_{oA} = e\int\psi_o \sum_i \vec{r}_i \psi_A^* d\tau$.

Associated with this movement of charge is a transition magnetic dipole moment,

(1-18) $\vec{m}_{Ao} = (e/2mc)\int\psi_A \sum_i (\vec{r}_i \times \vec{p}_i)\psi_o^* d\tau$.

In these equations ψ_o and ψ_A are the polymer wavefunctions for the ground and A^{th} excited state, $\vec{r}_i$ and $\vec{p}_i$ are the position and momentum vectors of the ith electron which has mass m and charge e. We can then define terms directly related to the absorption and circular dichroism for a particular absorption band.

(1-19) Dipole strength:

$$D_{oA} = \vec{\mu}_{oA} \cdot \vec{\mu}_{Ao} = (91.80 \times 10^{-40}) \times \int_{band}(\epsilon/\lambda)d\lambda$$

and

(1-20) Rotational strength:

$$R_{oA} = Im(\vec{\mu}_{oA} \cdot \vec{m}_{Ao}) =$$

$$(22.95 \times 10^{-40})\int_{band}[(\epsilon_L - \epsilon_R)/\lambda]d\lambda ,$$

where ϵ is the molar absorptivity, $\epsilon_L - \epsilon_R$ the circular dichroism in liters/(mole cm), λ is the wavelength in centimeters, and Im means the imaginary part. Despite the neat and simple appearance of these equations, which in principle relate molecular structure to electronic properties, the struggle still goes on -- not many experimenters have had to resign the field for lack of work. The wavefunctions presently available are not accurate enough to completely describe these

molecules.

We now proceed to more qualitative models whose approximate behavior can be easily described. More exact treatment of the general polymer case is contained in previous work.[42 - 49] Particularly useful for nucleic acids is the work on hypochromism by Devoe and Tinoco,[49] the work on exciton contributions to CD and UV absorption,[47] and a generalized optical theory by Tinoco.[42]

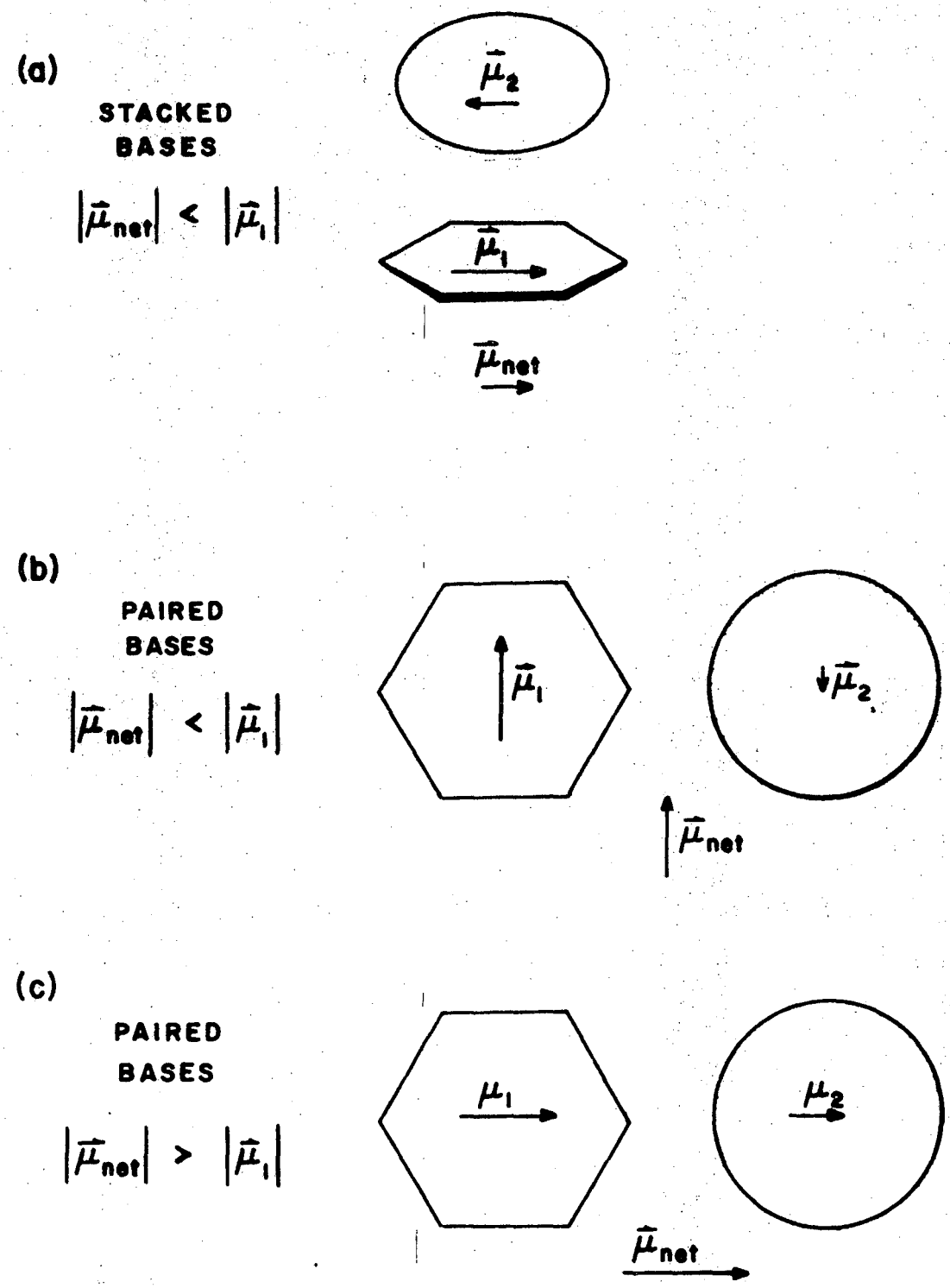
We can gain some insight into what to expect for hypochromism when bases are stacked or paired. First, consider a transition in a base with another base stacked on top. We can ascribe a polarizability to this second base that interacts with the transition moment of the first as in figure 1-4a. The transition $\vec{\mu}_1$ induces an opposing dipole, $\vec{\mu}_2$, in the polarizable group. The net polymer transition dipole is lower and the integrated intensity is lower. Therefore, we expect stacking to lead to hypochromism. For base pairing, either hypochromism or hyperchromism can result (see figure 1-4b,c). It is shown elsewhere[49] that in quantum mechanical terms hypochromism arises from the interaction between non-degenerate transitions on separate bases and is proportional to $(\vec{\mu}_{1a} \cdot \vec{\mu}_{2b})R_{12}^{-3}$ (a and b are non-degenerate excited states of monomers 1 and 2, and R_{12} is the distance between the transition moments). There can be a shift of intensity from one degenerate transition to another due to exciton splitting which can affect

0 0 0 0 3 9 0 0 8 6 5

Figure 1-4.

Interaction of a transition moment on one base with the polarizable electron cloud of another base. Base stacking (a) lowers the net transition dipole moment leading to hypochromism. Base pairing can lead to either hypochromism (b) or hyperchromism (c).

Figure 1-4


(a)
STACKED
BASES
$|\vec{\mu}_{net}| < |\vec{\mu}_1|$
$\vec{\mu}_2$
$\vec{\mu}_1$
$\vec{\mu}_{net}$
(b)
PAIRED
BASES
$|\vec{\mu}_{net}| < |\vec{\mu}_1|$
$\vec{\mu}_1$
$\vec{\mu}_2$
$\vec{\mu}_{net}$
(c)
PAIRED
BASES
$|\vec{\mu}_{net}| > |\vec{\mu}_1|$
μ_1
μ_2
$\vec{\mu}_{net}$

0 0 0 0 3 9 0 0 0 6 6

the band shape but not the integrated intensity.[47] It is important to remember that hypochromism is not a loss of intensity for the whole spectrum covering all frequencies. The sum of oscillator strengths (= constant · $\int_0^\infty \varepsilon d\nu$) must always equal the total number of electrons in the system. Intensity can be rearranged but not lost -- some of the shorter wavelength bands in polynucleotides must be hyperchromic. Finally we note that the angular dependence of hypochromism is contained only in the dot product $\vec{\mu}_{1a} \cdot \vec{\mu}_{2b}$.

We can further draw qualitative conclusions about CD. CD is related to $Im(\vec{\mu}_{oA} \cdot \vec{m}_{Ao})$. If an electron undergoes a transition in the presence of an asymmetric environment the transition is optically active. An asymmetric environment implies a distribution of permanent charges, dipoles, and polarizabilities in some combination which do not have mirror image symmetry. This means that an electron cannot move from a ground to excited state (actually the electron density changes) in a straight line path. If this path is not symmetrical, e.g. helical, a net magnetic moment results and the formal requirements for optical activity are met. Tinoco[42, 47] has developed equations for calculating the CD of nucleic acids. The equations are quite complex for polymers, involving interactions between all transitions on all of the bases. However, if we restrict ourselves to first-neighbor terms, we have a situation quite simi-

lar to the dimer. In any case, for electrically allowed transitions the large terms contributing to rotational strength are proportional to $\vec{R}_{12}\cdot\vec{\mu}_{1a}X\vec{\mu}_{2b}$. If we now examine the equations which are correct for the dimer case more thoroughly, the rotational strength contribution of the second transition interacting with the first will be

$$(1\text{-}21)\qquad R_{oa}(ob) = \left[\frac{\nu_a\nu_b}{\hbar c\left(\nu_b^2-\nu_a^2\right)}\vec{R}_{12}\cdot\vec{\mu}_{1a}X\vec{\mu}_{2b}\right]\times \left[R_{12}^{-3}\left(\vec{\mu}_{1a}\cdot\vec{\mu}_{2b}-\frac{3(\vec{R}_{12}\cdot\vec{\mu}_{1a})(\vec{R}_{12}\cdot\vec{\mu}_{2b})}{R_{12}^2}\right)\right]$$

where $a \neq b$. In the degenerate case, when $a = b$, two equal and opposite rotational strengths occur with a frequency splitting dependent on the second term in square brackets. As we can see, there is a much greater angular dependence for CD than for absorption and hypochromism because of the number and kind of vector products. CD should therefore be much more sensitive to geometry. Also CD has an R^{-2} dependence on distance while hypochromism has an R^{-3} dependence. This means that distant neighbors will contribute more to CD than to hypochromism. We can additionally make some interesting statements about the CD given specific geometries. The most favorable position of $\vec{R}_{12}$ for maximum rotational strength is when the transition moments are mutually perpendicular to it. Also, when the angle between tran-

0 0 0 0 3 9 0 0 8 6 7

sition moments is 45° in such a "stacked" conformation, the CD will be a maximum. When the transition moments are parallel the first factor in eq. (1-21) is zero, and when they are perpendicular the second term is zero. Also, coplanar transitions (e.g., in base pairs or other transitions in the same base) cannot contribute to the CD.

Equipped with the above generalizations about the influence of base stacking and pairing on CD we can predict what happens in real systems. We can say that as a helix of stacked bases (single or double) undergoes an order → disorder transition, the CD will decrease. We cannot say that a double helix will have a rotational strength higher or lower than its constituent single strands. The fact that it usually does implies that the double helix is more rigidly held in a stacked geometry. We must also say that all evidence so far accumulated indicates that neither RNA nor DNA has a stacked structure such as discussed above. Their geometries are close enough, however, that the qualitative conclusions are still valid. A final consideration on CD is that the sum of the measured rotational strengths for the region between 220 and 350 nm for RNA is usually far from zero while double strand "B-form" DNA has nearly equal positive and negative bands. The so-called "non-conservative" CD for RNA can be attributed to interactions between the transition moments in the neighborhood of 260 nm and far UV transitions.[49,50]

Interaction of the 260 nm transitions and static field terms are not large enough to account for the effect. This must mean that some far UV transitions are very intense (the factor $\nu_a\nu_b/(\nu_b^2 - \nu_a^2)$ is much smaller when the transitions are far apart in energy than when they are close). It must also mean that having the DNA bases in parallel planes perpendicular to the helix axis causes considerable cancellation of these interactions. Single strand DNA's can also give non-conservative spectra.[51]

iii. Other Methods

Many other physical methods have been used to obtain information about nucleic acid structure. It almost goes without saying that the greatest step toward understanding the molecular mechanism of nucleic acids came with the analysis of the x-ray diffraction patterns of DNA fibers. X-ray diffraction still stands as perhaps the ultimate in structure determination. However, one must first be able to crystallize the molecule of interest and then carry out a careful and quite complex analysis of the resulting patterns. On top of this, one must rely on other evidence to show that the structure present in the crystal form is the same as obtains <u>in vivo</u>.

Studies involving equilibrium dialysis are particularly applicable to situations involving the binding

of a small molecule such as an oligonucleotide to a larger polymer.[52,53] One particular advantage of equilibrium dialysis is its direct determination of equilibrium constants. There is not the necessity to construct theoretical relationships, for example, between hypochromicity and species concentration.

We could say much more about physical properties in general, but the ones described have a great deal of promise in structural elucidation for molecules such as the short oligonucleotides examined in this dissertation.

REFERENCES FOR CHAPTER I

1. Watson, J.D., Molecular Biology of the Gene, 2nd Ed., W.A. Benjamin, Inc., New York: 1970.

2a. Sanger, F., Brownlee, G.G., and Barrell, G.B., J. Mol. Biol., 13, 337 (1965).

2b. Brownlee, G.G. and Sanger, F., J. Mol. Biol. 23, 337, (1967).

3. Billeter, M.A., Dahlberg, J.E., Goodman, H.M., Hindley, J., and Weissman, C., J. Cell Phys., 74, No. 2 Sup., 197 (1969).

4. Tinoco, Jr., I., Levine, M.D., Uhlenbeck, O.C., Nature, 230, 362 (1971).

5. Delisi, C. and Crothers, D.M., Proc. Nat. Acad. Sci., U.S. 68, 2682 (1971).

6. Cramer, F., Progr. Nucl. Acid Res. Mol. Biol., 11, 391 (1971).

7. Gray, D.M. and Tinoco, Jr., I., Biopolymers, 9, 223 (1970).

8. Schildkraut, C., Marmur, J., and Doty, P., J. Mol. Biol., 4, 430 (1962).

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

10. Cantor, C.R., Jaskunas, S.R., and Tinoco, Jr., I., J. Mol. Biol., 20, 39 (1966).

11. Warshaw, M.M. and Cantor, C.R., Biopolymers, 9, 1079 (1971).

0 0 0 0 3 9 0 0 3 6 8

12. Warshaw, M.M. and Tinoco, Jr., I., J. Mol. Biol., 20, 29 (1966).

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

14. Gray, D.M., Tinoco, Jr., I., and Chamberlin, M.J., Biopolymers, 11, 1235 (1972).

15. Allen, F.S., Gray, D.M., Roberts, G.P., and Tinoco, Jr., I., Biopolymers, 11, 853 (1972).

16. Blum, A.D., Ph. D. thesis, U. of Calif. Berkeley, (1971).

17. Thomas, G., Biopolymers, 7, 325 (1969).

18. Tsuboi, M., Higuchi, S., Kyozoku, Y., and Nishimura, S., Biochim. Biophys. Acta, 195, 23 (1969).

19. Tsuboi, M., Takahashi, S., Muraishi, S., Kajiura, T., and Nishimura, S., Science, 174, 1142 (1971).

20. Kearns, D.R., Patel, D.J., and Shulman, R.G., Nature, 229, 338 (1971).

21. Kearns, D.R., Patel, D., Shulman, R.G., and Yamane, T., J. Mol. Biol., 61, 265 (1971).

22. Shoup, R.R., Miles, H.T., and Becker, E.D., Biochem. Biophys. Res. Comm., 23, 194 (1966).

23. Katz, L. and Penman, S., J. Mol. Biol., 15, 220 (1966).

24. Chan, S.I., Schweizer, M.P., Ts'o, P.O.P., and Helmkamp, G.K., J. Am. Chem. Soc., 86, 4182 (1965).

25. Schweizer, M.P., Chan, S.I., and Ts'o, P.O.P., J. Am. Chem. Soc., 87, 5241 (1965).

26. Broom, A.D., Schweizer, M.P., and Ts'o, P.O.P.,

J. Am. Chem. Soc., 89, 3612 (1967).

27. Chan, S.I., Bangerter, B.W., and Ts'o, P.O.P., Proc. Nat. Acad. Sci., U.S., 55, 720 (1966).

28. Davis, R.C., Ph. D. thesis, U. of Calif. Berkeley, (1967).

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

30. Ts'o, P.O.P., Kondo, N.S., Schweizer, M.P., and Hollis, D.P., Biochemistry, 8, 997 (1969).

31. Kondo, N.S., Holmes, H.M., Stempel, L.M., and Ts'o, P.O.P., Biochemistry, 9, 3479 (1970).

32. Fang, K.N., Kondo, N.S., Miller, P.S., and Ts'o, P.O.P., J. Am. Chem. Soc., 93, 6647 (1971).

33. Mantsch, H.H. and Smith, I.C.P., Biochem. Biophys. Res. Comm., 46, 808, (1972).

34. Imamoto, F. and Yanofsky, C., J. Mol. Biol., 28, 1 (1967).

35. Jones, A.J., Grant, D.M., Winkley, M.W., and Robins, R.K., J. Phys. Chem., 74, 2684 (1970).

36. Jones, A.J., Winkley, M.W., Grant, D.M., and Robins, R.K., Proc. Nat. Acad. Sci., U.S., 65, 27 (1970).

37. Jones, A.J., Grant, D.M., Winkley, M.W., and Robins, R.K., J. Am. Chem Soc., 92, 4079 (1970).

38. Tarpley, Jr., A.R. and Goldstein, J.H., J. Am. Chem. Soc., 93, 3573 (1971).

39. Dorman, D.G. and Roberts, J.D., Proc. Nat. Acad. Sci., U.S., 65, 19 (1970).

0 0 0 0 3 9 0 0 8 6 9

40. Tinoco, Jr., I. and Cantor, C.R. in Methods of Biochemical Analysis, Vol. 18, D. Glick ed., Wiley, New York (1970).

41. Warshaw, M.M., Ph. D. thesis, U. of Calif. Berkeley (1965).

42. Tinoco, Jr., I., Advan. Chem. Phys., 4, 113 (1962).

43. Condon, E.U., Rev. Mod. Phys., 9, 432 (1937).

44. Kirkwood, J.G., J. Chem. Phys., 25, 467 (1937).

45. Moffitt, W., J. Chem. Phys., 25, 467 (1956).

46. Moscowitz, A., Adv. Chem. Phys., 4, 67 (1962).

47. Tinoco, Jr., I., Radiation Res., 20, 133 (1963).

48. Tinoco, Jr., I., J. Am. Chem. Soc., 82, 4785 (1960).

48a. Tinoco, Jr., I., J. Am. Chem. Soc., 83, 5047 (1961).

49. Devoe, H. and Tinoco, Jr., I., J. Mol. Biol., 4, 518 (1962).

50. Bush, C.A. and Brahms, J., J. Chem. Phys., 46, 79 (1967).

51. Warshaw, M.M. and Cantor, C.R., Biopolymers, 9, 1079 (1970).

52. Uhlenbeck, O.C., Baller, J., and Doty, P., Nature, 225, 508 (1970).

53. Uhlenbeck, O.C., J. Mol. Biol., 65, 25 (1972).

CHAPTER II

Materials and Methods

A. General Considerations on Synthesis of Block Oligoribonucleotides

The preparation of oligoribonucleotides of precisely defined sequence is central to the studies described in this thesis. The molecules prepared are block copolymers, that is, sequences composed of short blocks of identical bases. Oligomers of complementary sequence prepared in this manner have a diminished tendency to form high molecular weight, multi-strand aggregates.[4] The aggregation problem has been frequently encountered by experimenters using simpler model compounds.[1,2,2a]

The techniques for synthesizing block oligoribonucleotides have been developed primarily by Thach[3] and further developed by Martin,[4,5] and Uhlenbeck.[4,6] These and other techniques are described below.

There is much to be said in favor of synthesizing molecules of self-complementary sequence. These usually have more complicated sequences, for example A_nGCU_n has the same kinds of interactions as a 1:1 molar mixture of $A_nGC + GCU_n$. Although the former requires three synthetic steps, three are also necessary for the

0 0 0 0 5 9 0 0 8 7 0

production of the latter pair. Multi-strand aggregation problems have never been encountered with self-complementary oligomers, whereas they have occasionally occurred with the mixtures. Finally, the self-complementary pair come pre-mixed, ready to add water. This is no small advantage since otherwise, one has to carefully determine extinction coefficients or do a mixing curve to get an exactly equimolar mixture. There is some advantage to being able to observe the behavior of the single strands independently, but the self-complementary molecules will dissociate at a low enough salt concentration.

B. Synthetic Reactions

Table 2-1 collects the data on the reactions used to make the oligomers in this study. Additional comments on the reactions are given in the sections below.

1. Partial Hydrolysis of Homopolymers

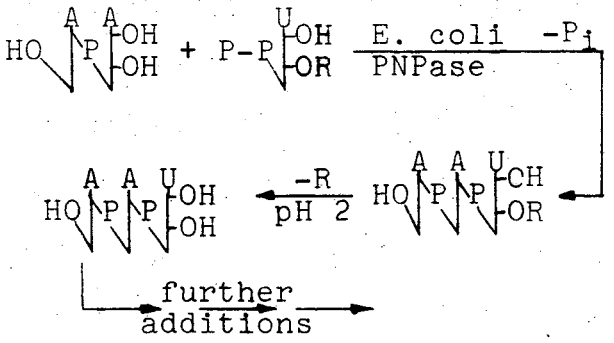
This method produces homo-oligomers, and is frequently used for the preparation of primers for the polynucleotide phosphorylase (PNPase) additions described below. Partial nuclease digestion of such polymers is frequently used as well as the base hydrolysis described here. The polymer solution should be heated to 60°C before adding base. Base hydrolysis is very fast and

Table 2-1. Useful reactions for oligo-RNA synthesis

Reaction:	Can prepare:	Reaction Conditions:	Comments:
Partial hydrolysis of homopolymers. e.g., poly A $\xrightarrow[\text{short time}]{OH^-}$ $A_{n-1}Ap$ $\xrightarrow{\text{terminal phosphatase}}$ $A_n + P_i$	A_n, C_n, U_n, short G_n	15 mg/ml poly X Heat to 60°C, add KOH to 1M. Quench reaction after 2 - 5 min. with $HClO_4$. Break cyclic phosphates at pH 2-3, 34°C, 2 hours.	<15% yield of maximum X_n, n>4. Do kinetic trials for proper chainlength distribution. G_n's aggregate.
Specific nuclease action on mixed copolymers. e.g., poly(A,C) $\xrightarrow[\text{exhaustive}]{\text{pan RNase}}$ A_nCp $\xrightarrow{\text{Ⓟ ase}}$ $A_nC + P_i$	A_nC, A_nU, G_nC, G_nU with pan RNase; A_nG, C_nG, U_nG with RNase T_1.	~ 3 µg RNase A/µmole C or U residue, or ~ 5 units RNase T_1/µmole G. 0.1 M Tris pH 8.2 0.01 M $MgCl_2$ ~ 6 µg Ⓟ ase/µmole terminal phosphate ~ 25 mg/ml polymer.	<15% yield of maximum A_nC, n>3. A/C ratio in polymer determines chain-length distribution.
Preparation of high molecular weight polymers. e.g., ADP + CDP $\xrightarrow{\text{PNPase}}$ poly(A,C) + P_i	Homopolymers or mixed co-polymers.	0.07 M NDP 0.2 M glycine-NaOH, pH 9.4 10 mM $MgCl_2$ 0.4-1 mg/ml PNPase.	<40% yield. Stop before extensive phosphorolysis. [ADP]/[CDP] ∝ distribution of A_nC sequences.

Primer dependent kinase reaction - simple block additions. e.g., A_n + UDP $\xrightarrow[0.8M\ NaCl]{PNPase}$ A_nU_m + P_i	Anything with a purine primer, low yields with pyrimidine primers (cannot make U_nA_m)	17.5 mM NDP 1-10 mM $MgCl_2$ 0.2 M glycine-NaOH pH 9.4 0.6-0.8 M NaCl 0.5-1 mM primer 0.2-2 P.U./ml PNPase (Tris-HCl buffer, pH 8.2 should be used if Ⓟase reaction follows.)	<15% primer incorporation into maximum A_nU_m, m<3. Do kinetic trials for each primer/NDP combination. [NDP]/[primer] ≈ 20:1 to limit phosphorolysis. Pyrimidines are poor primers. High salt concentration leads to good primer dependence, slower reaction. Long primers give better primer dependence than short ones.
Nuclease assisted kinase reaction. e.g., A_nC + GDP $\xrightarrow[0.4M\ NaCl]{PNPase}$ A_nCGp + P_i RNase T_1 >4 hours Ⓟase $\xleftarrow[2\ hrs]{pH\ 2,\ 37°C}$ boil 1-2 min $\rightarrow A_nCG$ + P_i	Any G terminal oligomer w/o internal G. With pan RNase, any C or U terminal oligomer w/o internal C or U.	35 mM GDP (CDP or UDP) 1-10 mM $MgCl_2$ 0.2 M Tris-HCl, pH 8.2 0.4 M NaCl 1-3 mM primer 300 $\frac{units}{ml}$ RNase T_1 or RNase A 1-2 P.U./ml PNPase 8-16 hours at 34°C.	80-100% incorporation of primer. Use as a trick to incorporate pyrimidine primers. 3'-P prevents phosphorolysis. Low salt leads to fast reaction. [GDP]/[A_nC] ≈ 20:1.

Nuclease assisted double kinase reaction. e.g., U_n + CDP + GDP $\xrightarrow[\text{0.4 M NaCl, RNase } T_1, \text{>4 hours}]{\text{PNPase}}$ $U_nC_mGp + P_i$ pH 2, boil → $U_nC_mG + P_i$ Ⓟase → $U_nC_mG + P_i$	G terminal oligomers w/o internal G, pyrimidine terminal oligomers w/o internal Py.	60 mM GDP (C or UDP) 0.5 mM primer otherwise identical to above reaction incubate 8-16 hr at 34°C.	Good primer incorporation. Must do [CDP]/[GDP] trials to get proper distribution - also check to see that each oligomer ends in G. [NDP]/[primer] ≈ 100:1.
Kinase/phosphorylase equilibrium reaction. e.g., GC + UDP $\xrightarrow[\text{0.4 M NaCl}]{\text{PNPase}}$ GCU_{100} + P_i $\xrightarrow[\text{(fast)}]{\text{kinase}}$ equilibrium ~40 hours ← GCU_{99} + UDP ← phosphorylase (slow) ↓ → $GCU_n + P_i$	Dimer-X_n where $X \neq G$.	$7 \cdot n_{avg}$ mM NDP 7 mM dimer 0.4 M NaCl 0.2 M glycine-NaOH pH 9.4 1-2 P.U./ml PNPase (Tris-HCl buffer pH 8.2 should be used if Ⓟase reaction follows) Incubate >36 hours at 34°C.	<20% yield of maximum GCU_n, $n<4$. Phosphorolysis starts at 3' end and proceeds only to dimer. Dimers are cheap. Distribution determined by [UDP]/[GC] - should do trials but can often guess close enough.

Gilham reaction - blocked diphosphates. e.g.,  	Any sequence!	See reference 13.	R blocks polymerization at 3'-OH. Works only with E. coli PNPase. Works better with longer primers. Additions not quantitative, yield ~ 70% adding ADP, ~ 50% adding CDP or UDP.

therefore somewhat difficult to control. However, the ease with which all "enzyme" activity is destroyed by simply adding acid makes it the method of choice. Using KOH for the hydrolysis and $HClO_4$ for the neutralization has the advantage of producing precipitated $KClO_4$, thus keeping the final salt concentration low. This means no further dilution is necessary before applying to an anion exchange column. After neutralizing, the pH should be adjusted to 2 and the solution incubated for two hours at 37°C to break cyclic phosphates. If only one or two short oligomers are the desired products the removal of the terminal phosphate may be postponed until after the column separation, with a saving in cost of purified alkaline phosphatase. We have used bacterial alkaline phosphatase (Worthington Biochemicals, grade BAPC) dialyzed into 10mM $MgCl_2$, 0.1 M NaCl, 0.01 M tris (pH 8). This and the cheaper grade (BAPSF) contain a contaminating endonuclease with very interesting properties. This so-called RNaseAc cleaves after cytidine residues producing an oligomer ending in cytidylyl 2'-3' cyclic phosphate.[33] RNaseAc cuts poly C into oligomers but degrades them to monomers only very slowly. The dimers CpC, UpC, ApC, and GpC are not appreciably degraded on exhaustive treatment.[15] We have also determined that BAPC preparations degrade CpU, CpC, and CpG less than 5% (incubated 20 hours at 37°C with 1 mg/ml BAPC phosphatase, RNaseAc concentration undetermined,

0 0 0 0 5 9 0 0 8 7 3

0.1 M tris pH 8, 0.4 M NaCl, 10 mM $MgCl_2$, and 8 mM CpX). CpA, however, under the same conditions is approximately 40 % degraded. We have also observed less than 5% activity on CpC, CpU, and CpG at a lower salt concentration (1 mM $MgCl_2$, 10 mM NaCl) and a reduced, ~25 %, activity on CpA. No attempt has been made as yet to rigorously determine the sequence dependent properties of a purified preparation of RNaseAc but from the results above it could become a powerful sequencing tool. It has also been our experience that CAG is similarly degraded, so the use of RNase free phosphatase (eg, Worthington BAPF) or avoiding its use entirely is advised for such preparations.

2. Specific Nuclease Action on Mixed Copolymers

The products of this reaction are also frequently used starting points for PNPase additions. The nuclease should be removed before application to the anion exchange column, since nuclease activity commonly co-chromatographs with the trimer through pentamer peaks on a neutral Tener-Tomlinson column. The nuclease may be removed by a double phenol extraction followed by dialysis to get rid of the dissolved phenol. We typically dialyze against 0.1 M NaCl, 0.01 M tris (pH 8) overnight. An alternative procedure is to treat twice with bentonite[7] (1 mg/mg of nuclease is sufficient), centrifuging out the bentonite-nuclease complex. Again, the

terminal phosphate removal may be postponed until after the initial separation is effected. The phosphatase can be inactivated by heating above 90°C for one or two minutes. Also, the phosphatase and nuclease steps can be carried out at the same time. The nucleolytic hydrolysis should be carried to completion to insure that all bases are hydrolyzed for which the enzyme is specific.

3. Polynucleotide Phosphorylase Reactions

The following reactions involve use of the enzyme polynucleotide phosphorylase. This enzyme catalyzes a kinase reaction of a mononucleoside-5'-diphosphate to the 3' end of a growing polynucleotide, with the release of inorganic phosphate. It also catalyzes the reverse or phosphorylase reaction. Two different forms of the enzyme are known for several organisms, primer dependent and primer independent polynucleotide phosphorylase. Primer dependent enzyme is characterized by a lag period in synthesis which does not occur if a "primer" oligonucleotide with a free 3'-hydroxyl is added. The primer must be a dinucleoside monophosphate or longer. The ratio of primed to non-primed synthesis is defined as the "primer dependence". The two forms of the M. luteus enzyme have different electrophoretic properties[8] with the latter convertible to the former by a partial trypsin digest. Singer[9] has described a procedure to prepare primer dependent enzyme, but its success appears

0 0 0 0 3 9 0 0 3 7 4

to depend on the particular batch of cells from which the crude enzyme is obtained. Crude (primer independent) enzyme can be obtained from P. L. Biochemicals, is quite inexpensive, and is suitable for the preparation of high moelcular weight polymers. Occasionally, chromatographing this crude enzyme on a Sephadex (Pharmacia Chemicals) G-200 molecular sieve column as described by Klee[8] will generate primer dependent enzyme without trypsin treatment. The rapid and reproducible application of G-200 columns by using an internal support of glass beads to avoid gel compression has recently been introduced[10], so routine scanning of several lots of crude PNPase for primer dependence is now possible. P. L. Biochemicals also sells primer dependent enzyme although it is presently quite expensive. The trypsin treatment destroys some of the enzyme activity and is occasionally unsuccessful in generating sufficient primer dependence. However, starting with the crude commercially available enzyme avoids the most difficult and unpredictable stage of the preparation. We have also found that PNPase can be concentrated by lyophilization with a minimal loss of activity. For more information see the appendix to Chapter II on assaying PNPase.

a. Preparation of High Molecular Weight Polymers

Table 2-1 gives the conditions for the primer independent kinase reaction used to synthesize long poly-

nucleotides. Occasionally, even "primer independent" enzyme displays a lag in synthesis (although it is not primer dependent enough to be used as such). Therefore, it is frequently helpful to add a small amount of a dimer compatible with the base composition being synthesized. The course of the reaction may be followed by colorimetrically monitoring inorganic phosphate release, or by running a small parallel reaction with one radioactive nucleoside diphosphate and following TCA precipitable counts. The reaction should be stopped when the reaction rate begins to slow. This usually corresponds to something less than 50% yield. The reason that the rate of incorporation slows is that sufficient inorganic phosphate has built up for the phosphorolysis to proceed at a reasonable rate. If the reaction is allowed to proceed much past this point the products of the reaction will be short polymers. This is fine for synthesis of homopolymers used for the subsequent preparation of homo-oligomers, but for mixed co-oligomers made by a specific nuclease digestion, such as the series A_nC, a considerable fraction of the oligomers will not end in C. To halt the reaction, the enzyme can be denatured by heating above 90°C for one to two minutes. For the preparation of poly G and copolymers with more than about a third G, the divalent cation Mn^{2+} is preferred over Mg^{2+} [11].

The distribution of sequences within the random

0 0 0 0 3 9 0 0 3 7 3

copolymer is proportional to the ratio of nucleoside diphosphates used in the preparation. The proportion is not exact, however, since the enzyme prefers some substrates to others. The exact nature of this preference has not been ascertained. However, in one preparation of poly (AC) in this work, the ADP/CDP ratio in the reaction mixture was 4/1 while the distribution centered roughly around A_3C. Another general tendency is that G residues tend to accumulate in blocks[12]. A useful relation for the nuclease products of a generalized polymer, poly (XY), is that the mole fraction of bases in a sequence X_nY is:

$$f_{X_nY} \simeq \frac{nR^n}{(R+1)^{n+1}}$$

where

$$R = \frac{[XDP]}{[YDP]}$$

Then one should get approximately

$$\frac{(n+1)R^n}{(R+1)^{n+2}} \times \text{(number of Absorbance units of polymer)}$$

absorbance units of the oligomer X_nY. This formula ignores differences in the extinction coefficients of X and Y, and assumes the bases are inserted randomly according to the ratio of diphosphates used in the preparation of the polymer.

b. Primer Dependent Kinase Reactions -- Simple Block Additions

This type of reaction is used for the addition of several identical bases to a primer molecule leading to a distribution of oligomers of various lengths. The parameters affecting chainlength have been selected so that the only necessary variables are time of incubation and the PNPase concentration. However, each NDP and primer behave differently, so small-scale trial reactions must be performed to determine the optimal time and enzyme concentration.

In order to clarify the following discussion it is necessary to define the terms primer dependence (P.D.), and polymerization units (P.U.). P.D. is the ratio of primed to non-primed synthesis and varies considerably with the enzyme preparation. Polymerization units are defined in terms of the initial rate at which ADP is added to oligo A_4 primer, under standard conditions. This assay is described in the appendix to Chapter II.

Generally, the primer dependent kinase reaction as described here is carried out at a NaCl concentration of 0.8 M. For oligomers nearing twenty bases long it is better to run at a concentration of 0.6 M since the oligomer length appears to be limited by salt concentration[1]. The choice of this high salt concentration is made because the primer dependence increases with salt concentration. Also, Uhlenbeck[6] and Martin[5] have

0 0 0 0 3 9 0 0 0 7 6

extensively characterized the oligomer synthesis process at this salt concentration. Of course, it should be pointed out, that synthesis can be carried out over a very wide range of salt concentration[1].

The optimal Mg^{2+} concentration varies with each preparation of PNPase and should be determined before a synthesis program is started. The optimal rate of synthesis usually occurs somewhere between 1 and 10 mM. It also appears[1] that the Mg^{2+} concentration may limit the length of oligomers that can be synthesized.

Another general characteristic of this enzyme system is that purine oligomers are better primers than pyrimidines. Oligo U's are abysmally poor primers, and cannot comprise the majority of the primer unless one of the nuclease tricks (discussed in later sections of this thesis) are used. For instance, the standard reaction to produce U_nA_m by adding ADP to U_n goes in very poor yield because as soon as one A is added to a few of the primers, the U_nA produced is a much better primer than U_n. The result is a few very long U_nA_m's and a lot of leftover U_n[5,6]. We have been able to synthesize U_2CGA_2 in low yield, working at very high enzyme concentration so more than a usual number of synthesis starts are made on many primers.

Factors that influence primer dependence include the salt concentration, as was already mentioned, the primer length, and the particular NDP. Longer primers

generally lead to a faster rate of polymerization than shorter ones, all else being equal. ADP tends to self-prime more than the other NDP's.

The NDP/primer ratio is kept high, usually about 20/1, to keep phosphorolysis from becoming a serious problem. Phosphorolysis would lead to randomization of the sequence of the oligomers produced by disassembling the primer and mixing in the diphosphates removed with those intended to go on the 3' end of the primer.

The proper enzyme concentration is determined largely by the length of the primer since the rate of polymerization is greater for long primers. With primers of length two or three, we use enzyme concentrations of 0.8 to 1.0 polymerization units/ml. Generally, for longer primers 0.2 to 0.4 units/ml is satisfactory. Also, we use high enzyme concentrations for additions of long blocks and with pyrimidine rich primers. The object of adjusting the enzyme concentration is largely to permit the reaction to take place in a reasonable length of time -- 1/2 to four hours. This usually makes it easier to determine the proper time of incubation, than for a very short or very long reaction time. A sample time trial and synthesis of A_4U_m is included in the appendix to Chapter II.

0 0 0 0 3 9 0 0 3 7 7

c. Nuclease Assisted Kinase Reaction

The second major class of primer dependent PNPase reactions involves coupling the addition reaction with a nuclease specific for the residue being added. This allows the nearly quantitative addition of one base to a primer. (See Table 2-1 for an illustration of the application of this reaction.) The reaction conditions are modified somewhat from those of the previous reaction, viz., the pH is lowered to 8.2 since both RNase T_1 and pancreatic RNase (hereafter called RNase A) don't work well at 9.4; the salt concentration is lowered to 0.4 M to increase the rate of synthesis (self-priming is not a problem since the resulting oligomers will be hydrolysed to monomers); and the reaction time lengthened to insure complete reaction. The 3'-terminal phosphate left by the nuclease action prevents phosphorolysis of the oligonucleotide so the time of incubation may be quite long. The NDP/primer ratio is still kept high to drive the reaction in the direction of products since the RNases also catalyze the reverse reaction. This sort of use of RNase T_1 is a convenient method for the incorporation of pyrimidine rich primers which would otherwise result in very low yields.

d. Nuclease Assisted Double Kinase Reaction

Looking at the example in Table 2-1 we see that

this is really only a slight variation of the above reaction. The advantage is the specific addition of two blocks of residues at the same time with good yields of products. The conditions are identical to the previous reaction with the exception of an increase in the NDP to primer ratio. This is done simply because more than one residue is being added to the primer. Without this precaution one frequently finds that the molecules produced do not all have the same base at the 3'-terminus. The suggested ratio of 100/1 should probably be increased if the desired distribution is to center around m = 4 or longer, e.g., to make U_3C_5G from U_3,CDP, and GDP. Trials with varying ratios of diphosphates should be carried out to achieve the proper distribution and also to verify that each molecule ends in the desired base. This latter goal may be accomplished by a double label experiment or parallel trials with one radioactive diphosphate and then the other. A possible source of difficulty in a synthesis such as U_3C_nG is the production of oligo C_k by T_1 --

$$U_3 + CDP + GDP \xrightarrow{PNPase} U_3C_nG_jC_k \xrightarrow{T_1} U_3C_nGp + (j-1)Gp + C_k.$$

This oligo C_k may in fact be a much better primer than the U_3 making the separation of products very difficult.

0 0 0 0 3 9 0 0 8 7 8

e. Kinase/phosphorylase Equilibrium Reaction

This is the last major class of primer dependent PNPase reactions and is also illustrated in Table 2-1. In the syntheses up to now the phosphorylase action has been intentionally subdued, but here we take advantage of the fact that it proceeds only to the dinucleoside. The kinase activity adds nucleotides to the dimer and also produces inorganic phosphorous. Large scale phosphorolysis starts at the 3' end after a sufficient pool of inorganic phosphate has accumulated. The addition and excision reactions continue and finally reach a state of equilibrium which depends on the initial NDP to dimer ratio. The final distribution is proportional to this ratio, but rarely is the correlation exact. Therefore, very large scale preparations demand a previous trial. On a smaller scale, however, since dimers and NDP's are quite inexpensive, the distribution can be guessed at and usually enough of the desired oligomers are produced. Low yields are the rule for producing G adducts in this manner since they precipitate when they grow longer than 4 or 5. It may be possible to carry out this reaction using Mn^{2+} as the counterion as in the production of poly G[11].

This concludes our discussion of the synthetic techniques used in preparing the oligomers for this study. Table 2-2 summarizes the syntheses.

Table 2-2

		REACTION CONDITIONS								YIELD			
Oligomer	Source	P.U.[a] / ml	Primer	Time (min)	[NaCl]	[NDP] / [Primer]	pH	RNase	RNase units / ml	A_{260} units	%[b]	Trial reaction[c]	Comments[c]
A_4U_4	this work	0.34	A_4U^d	90	0.8	25	9.4	--	---	26	15	^{3}H-UDP, separate A_4U_n in 60 E ~4 days.	Distribution centers on $A_4U_{2.5}$, 120 min better.
A_5U_5	"	"	A_5U^d	120	"	25	"	--	---	4	2	Guessed (badly) from A_4U_n and A_4GU_n preps.	Distribution centers on $A_5U_{2.5}$, 240 min better. Acid column.
A_6U_6	"	1.3	A_6^e	300	"	18	"	--	---	35	22	^{3}H-UDP, after rxn add RNase A, count Up/U ratio after sep. in 70 E.	Probably salt concentration too high. Acid column. Good distribution.
A_4UA	"	0.22	A_4U^d	50	"	25	"	--	---	100	24	^{3}H-UDF, sep. A_4UA_n in 50 E, ~2 days.	Distribution centers on $A_4UA_{0.4}$, 80 min better.
A_4UAU_4	"	1.1	A_4UA	240	"	23	"	--	---	20	12	^{3}H-UDP, run off excess hot UDP in 70 E, elute, add RNase A, count Up/U ratio after sep in 70 E.	Good distribution.
A_4GU_3	"	0.22	A_4G	210	"	25	8.2	--	---	30	12	^{3}UDP, boil time aliquots, add T_1 and Ⓟase, optimize ratios of oligo U's.	Acid column. Distribution centers on A_4GU_6.
A_4GU_5	"	"	"	"	"	"	"	"	"	26	8		
A_3CU_4	"	0.68	A_3C	100	"	20	9.4	--	---	24	7	^{3}H-UDP, sep A_3CU_n in 70 E, 5 days	Distribution centers on A_3CU_2, 150 min better.
A_5CU_4	"	0.34	A_5C	180	0.8	22	"	--	---	20	10	Guessed from A_5U_n prep.	Dist centers on A_5CU_2 120 min better.

A_2GC	"	0.90	A_2G^e	300	0.4	18	8.2	A	300	150	60	none	Sep products in 60 E ~18 hr. Ⓟase after eluting, restreak in 40 E, ~18 hr.
A_2GCU_2	"	0.45	A_2GC	50	0.8	23	9.4	--	---	64	27	^{3}H-UDP, sep A_2GCU_n in 60 E ~3 days.	Perfect!
A_3GC	"	0.90	A_3G^e	180	0.4	46	8.2	A	300	76	60	none	Same as for A_2GC prep.
A_3GCU_3	"	0.90	A_3GC	85	0.8	27	9.4	--	---	10	8	Same as for A_2GCU_2	Dist centered on $A_3GCU_{2.5}$, lost a lot due to troubles with TEAE-cellulose column. 3rd column (DEAE) okay.
A_4GCU_4	OCU[e]												
A_5GC	this work	1.12	A_5G^e	180	0.4	60	8.2	A	300	90	62	none	Sep on column. Restreak in 30 E.
A_5GCU_5	"	0.45	A_5GC	155	0.8	24	9.4	--	---	12	8	^{3}H-UDP, add T_1 to time aliquots, sep in 60 E, count Up/U ratio.	Distribution centered on A_5GCU_3, 200 min better.
A_2CG	"	0.90	A_2C	300	0.4	15	8.2	T_1	250	260	90	none	Sep products in 40 E ~18 hr. Ⓟase after eluting, restreak in 40 E ~18 hr.
A_2CGU_2	"	0.45	A_2CG	80	0.8	23	9.4	--	---	24	13	Trial done same as for A_2GCU_2 prep.	Dist centers on $A_2CGU_{3.5}$, 50 min better.
A_3CG	"	0.26	A_3C	300	0.4	17	8.2	T_1	250	190	95	none	Same as in A_2CG prep.
A_3CGU_3	"	0.34	A_3CG	105	0.8	36/20	9.4	--	---	5	6	Guess from A_3GCU_3	Ran rxn two times. 2nd time NDP/primer too high. Dist centered at $A_3CGU_{1.5}$ 1st time looked okay but peak lost due to fraction collector failure.
A_4GC	"	0.26	A_4G	240	0.4	35	8.2	A	300	50	80	none	Sep A_4GCp on column, Ⓟase, step gradient on 3 ml column.
GCU_4	OCU[e]												

A_4CG	this work	0.90	A_4C	300	0.4	19	8.2	T_1	500	24	42	none	Sep A_4CGp from GDP in 40 E, Ⓟase, run on column to sep A_4CG and A_4C.
CGU_4	"	0.90	CG	48 hr	0.4	3.5	8.2	--	---	25	2	none	Should increase NDP/ primer to 5 or 6.
A_3CU	"	0.68	A_3C	90	0.8	20	9.4	--	---	40	18	Used trial from previous A_3CU_n.	Dist centered on A_3CU_2, 50 min better.
A_3CUG	"	0.26	A_3CU	8 hr	0.4	36	8.2	T_1	250	45	95	none	Sep on column.
CAG	"	2.2	CA	18 hr	0.4	48	8.2	T_1	300	102	37	none	Boil, Ⓟase, and streak in 30 E. BAFC Ⓟase has contaminant that hydrolyzes CAG → Cp cyclic + ApG.
$CAGU_3$	"	0.82	CAG	100	0.8	18	8.2	--	---	34	17	^{3}H-UDP, sep $CAGU_n$ in 70 E ~4 days.	Dist centered on $CAGU_4$.
$CAGU_5$	"	"	"	"	"	"	"	"	"	20	10	"	"
A_4G_2	OCU[e]												
C_2U_4	"												
U_2C	this work	2.2	UpU	50 hr	0.4	1.5	8.2	--	---	110	17	none	Ⓟase, separate on paper in 70 E.
U_2CG	"	2.2	U_2C	18 hr	0.4	32	8.2	T_1	300	92	66	none	Ⓟase, sep on paper in 60 E. Elute and restreak in 60 E.
U_2CGA_2	"	2.2	U_2CG	30	0.8	23	9.4	--	---	8	6	^{3}H-ADP, sep U_2CGA_n in 60 E, ~3 days.	Many long adducts. Sep on column. A real triumph!

[a] Polymerization units in 0.8 M NaCl.

[b] % incorporation of primer into product.

[c] Numbers such as "60 E" indicate paper chromatography in "60" % v/v (95% ethanol), 40%(1 M $NH_4C_2H_3O_2$).

[d] Gift from Mr. Eric Wickstrom.

[e] Gift from Dr. O. C. Uhlenbeck.

0 0 0 0 3 9 0 0 3 8 0

f. Gilham Reaction -- 3'-blocked Diphosphates

If a normal nucleoside-5'-diphosphate has a blocking group on the 3'-hydroxyl, it can be added to the 3' end of a primer as has been the case in the PNPase reactions described above. However, no further polymerization can take place until the 3'-block is removed. This essentially describes the recent and powerful technique devised by Gilham[13], which should allow the synthesis of RNA of any sequence in good yield. The method is illustrated schematically in Table 2-1 and we include it only because of its potential importance. None of the oligomers studied in this work were synthesized by the Gilham technique. Several blocking groups satisfy the requirements of being acid-labile and being small enough to mimic the behavior of a normal diphosphate with respect to PNPase. Gilham seems to prefer a methoxyethyl substituent which can be removed by incubating at pH 2 for 15 minutes, and is incorporated in good yield. The 2',3'-disubstituted diphosphate is not incorporated by the enzyme. Uhlenbeck[12] reports that PNPase isolated from E. coli gives much better yields than does M. luteus enzyme. He also has done experiments which show that A is added in higher yield than C and U. The latter are incorporated about 50% of theoretical. A further result is that there is a definite primer dependence; a blocked diphosphate is added much more readily to long primers

than to dimers. It seems likely that reactions similar to the nuclease assisted double kinase reactions described in Chapter II.B.3.d. should be possible. For instance, one should be able to carry out these syntheses:

$$A_3 + ppU + ppC^* \xrightarrow[PNPase]{E.coli} A_3U_nC^* \xrightarrow{pH2} A_3U_nC$$

or

$$U_4 + ppA + ppA^* \xrightarrow[PNPase]{E.coli} U_4A_{n-1}A^* \xrightarrow{pH2} U_4A_n$$

where ppN is a normal diphosphate and ppN^* is the 3'-blocked species. Such reactions, when applicable, would save several steps over the previously described Gilham reactions.

g. Carbodiimide Reaction -- Blocking Nuclease Action

Gilham and co-workers[17,18,19] have developed a method for blocking the action of ribonuclease by substituting the uridines and guanines in an intact oligonucleotide with N-cyclohexyl-N'-β-(4-methylmorpholinium)-ethylcarbodiimide p-toluenesulfonate. This carbodiimide blocking agent can be removed by incubation at pH 10.5 for 20 hours at 20°C. This is a useful method for producing the following: U_nC (from random poly UC) and oligomers containing internal U's (for example, AUC from AUC_n). The use of this technique is complicated when the blocking group must be added to an oligomer of considerable secondary structure[15].

0 0 0 0 3 9 0 0 0 8 1

4. Other Nuclease Reactions

The production of more complex sequences with the primer dependent PNPase technique would be greatly facilitated if one had a purine or adenine specific nuclease that could operate under the same conditions of salt and pH as does PNPase. The use of pancreatic RNase and RNase T_1 in such a manner has been explained in sections B.3.c. and d. of this chapter. There does exist a reasonably A specific nuclease in RNase U_2[20]. It has a pH optimum of 4.5, however, so it cannot be used directly with PNPase. It also has a tendency to leave some adjacent A's uncleaved, as well as cutting after some of the G's[20]. It would probably be suitable for the production of some oligomers, for example U_nA could be made from random poly AU. The small contaminating amounts of $U_{n-1}A_2$ in the U_nA from a neutral pH column could be removed by running a second acid column (described in section C below).

The use of RNase T_1 as a joining enzyme has been attempted with limited success[5,22]. First, the 3' phosphate of an oligomer XpYpZpGp must be reacted with ethyl chloroformate to form the 2'-3' cyclic phosphate. Then it is joined at very high concentration to the oligomer of choice with a free 5'-OH, in the presence of the enzyme. More details of this reaction are given in the references cited above.

C. Separation Techniques

The primary separation methods used in oligo-RNA synthesis are simple one-dimensional paper chromatography and anion exchange column chromatography. In paper chromatography we have used a simple solvent system of varying parts of 95% ethanol and 1M ammonium acetate at approximately neutral pH. When one increases the ethanolic component, mobility decreases but the resolution is better. The primary aim is to get as good a separation as is possible in a minimum length of time. A striking effect in paper chromatography is the much higher mobility of pyrimidines relative to purines. This is demonstrated by the data in Table 2-3 taken from Thach[3]. The relative mobilities clearly indicate that a separation of the homologous series, A_3U_n is better attempted in a high ethanol solvent since the separation is based on the number of pyrimidines, while a separation of U_3A_n should be carried out in a high aqueous solvent. We have used solvent compositions varying from 70% aqueous for quick separations of monomers from primer and products, to 30% aqueous with running times up to a week when high resolution is necessary. A typical chromatogram is shown in the appendix to Chapter II.

The amount of material in a 1 cm diameter spot should not exceed about 15 A_{260} units (less for oligomers

0 0 0 0 3 9 0 0 8 0 2

Table 2-3. Effect of Solvent Composition on Chromatographic Mobility of Oligonucleotides.[a]

Oligomer	Distance Migrated (cm) 55 E solvent (20 hr)	30 E solvent (16 hr)
UpU	31.2	37.7[b]
$(Up)_3U$	23.0	35.0[b]
U_6	16.3	31.8[b]
U_8	9.2[b]	27.6[b]
U_{10}	5.5[b]	25.0[b]
UDP	25.5	--
CpC	29.4	35.0[b]
C_4	18.5	30.2[b]
C_6	9.9	25.8[b]
C_8	5.1[b]	21.6[b]
C_{10}	2.4[b]	--
CDP	22.2	36.7
ApA	23.0	25.3
A_3	--	19.6
A_4	8.6	13.4
A_5	4.2	8.3
A_6	1.6[b]	5.0[b]
A_7	0.8[b]	2.5[b]
ADP	20.0	--

[a]Taken from ref. 3.

[b]Incompletely resolved from shorter or longer chains.

with blocks of G) or overloading artifacts are observed. Oligonucleotides may also be "streaked" along about 10 to 15 cm of a 20 cm wide paper for purifying preparative amounts (no more than 200 A_{260} units/15 cm, and less than about 0.1 mmole NaCl/15 cm)[3]. After development the strip containing the oligomer of interest is then dried, washed 2 - 3 times in absolute ethanol, dried, and eluted in water or buffer. The elution is best carried out in a closed container since the solvent front moves faster in a vapor saturated atmosphere. The streaking technique is not useful for preparing samples for UV spectroscopy since UV absorbing material also elutes from the paper.

The other principal separation technique we have used is anion exchange column chromatography. We have used DEAE-cellulose (Bio-Rad Laboratories Cellex-D, standard capacity) and DEAE-sephadex (Pharmacia Fine Chemical Co. grade A-25) as the anion exchanger. (DEAE stands for diethylaminoethyl.) We have used mainly the Tener-Tomlinson type column[14] -- 7 M urea to destroy secondary structure, a linear salt gradient from 0 to 0.275 M NaCl (higher if the oligomers are quite long) for DEAE-cellulose, or from 0.05 to 0.45 M NaCl for DEAE-sephadex, and either at pH 8.2 in 0.01 M tris or at pH 3.0 - 3.5 in 0.01 M sodium formate. The effect of pH on the separation is discussed below.

The oligomer length limit for good resolution

0 0 0 0 3 9 0 0 8 3

depends on the molar amount of phosphate, the distribution of oligomers, column size, flow rate, gradient shape, pH, and base composition. The following general guidelines are based on experience and, while perhaps not the ideal situation, are satisfactory for the synthesis of oligomers such as those studied here. More rigorous discussions of ion exchange theory can be found elsewhere[15,16]. The amount of phosphate should be low enough that it can all stick to the top ten percent of the column. The available ion exchange capacity for oligo A_5 with no terminal phosphates has been determined to be approximately 50 A_{260} units per ml of packed gel for the DEAE-dellulose, and 200 A_{260} units per ml for DEAE-sephadex. Both of these determinations were made according to the following procedure: A_5 in 7 M urea at pH 8.2 and approximately 0.05 M NaCl was loaded on 2 ml gel columns. When the OD of the effluent began to rise, the A_5 solution was diluted with 7 M urea to a NaCl concentration of about 0.02 M. When the OD of the effluent again rose the column was washed with distilled water, 0.01 tris pH 8.2 until A_{260} fell to zero. The bound A_5 was bombed off with about 30 ml of 0.5 M NaCl, 0.01 tris pH 8.2, until the OD of the effluent again fell to zero. The number of A_{260} units was calculated and divided by the column volume to get the available capacity for A_5. These numbers are probably the most useful form to express the available capacity for oligo-

nucleotide synthesis, but if the average of the distribution is much different from N_5 an available capacity figure in terms of moles of phosphate would be more helpful. Thus, for DEAE-cellulose the capacity is about 3.3 µmole phosphate/ml of packed gel, and for DEAE-sephadex, about 13.2 µmole internal phosphate/ml of packed gel. Diphosphates can be treated roughly the same as tetranucleotide triphosphates since they each have three negative charges. It has been observed that A_4 appears as a shoulder at the end of the ADP peak on a DEAE-cellulose column, pH 8, 7 M urea, with a linear salt gradient.[12]

The distribution of oligomer lengths influences the quality of anion exchange separation in that with a linear gradient the longer oligomers run closer together than the shorter ones. Thus, a shallower gradient and/or a slower flow rate is necessary to effect a good separation of a large quantity of long oligomers. If one still has problems with a poor separation such as is illustrated in figure 2-1a, the single oligomer of interest can be rechromatographed, or the entire series may be purified on two additional columns. One column with the odd numbered peaks, as in figure 2-1b, and the other with the even numbered peaks, figure 2-1c. To save time, since the individual peaks from the first column have a large volume, the odd peaks should be concentrated by diluting them with 3 volumes distilled water, then

0 0 0 0 3 9 0 0 0 0 4

Figure 2-1.

Hypothetical enhancement of separation of poorly resolved oligomers by rechromatography on two columns. The enhanced peaks may now be cut at the arrows to give highly purified fractions.

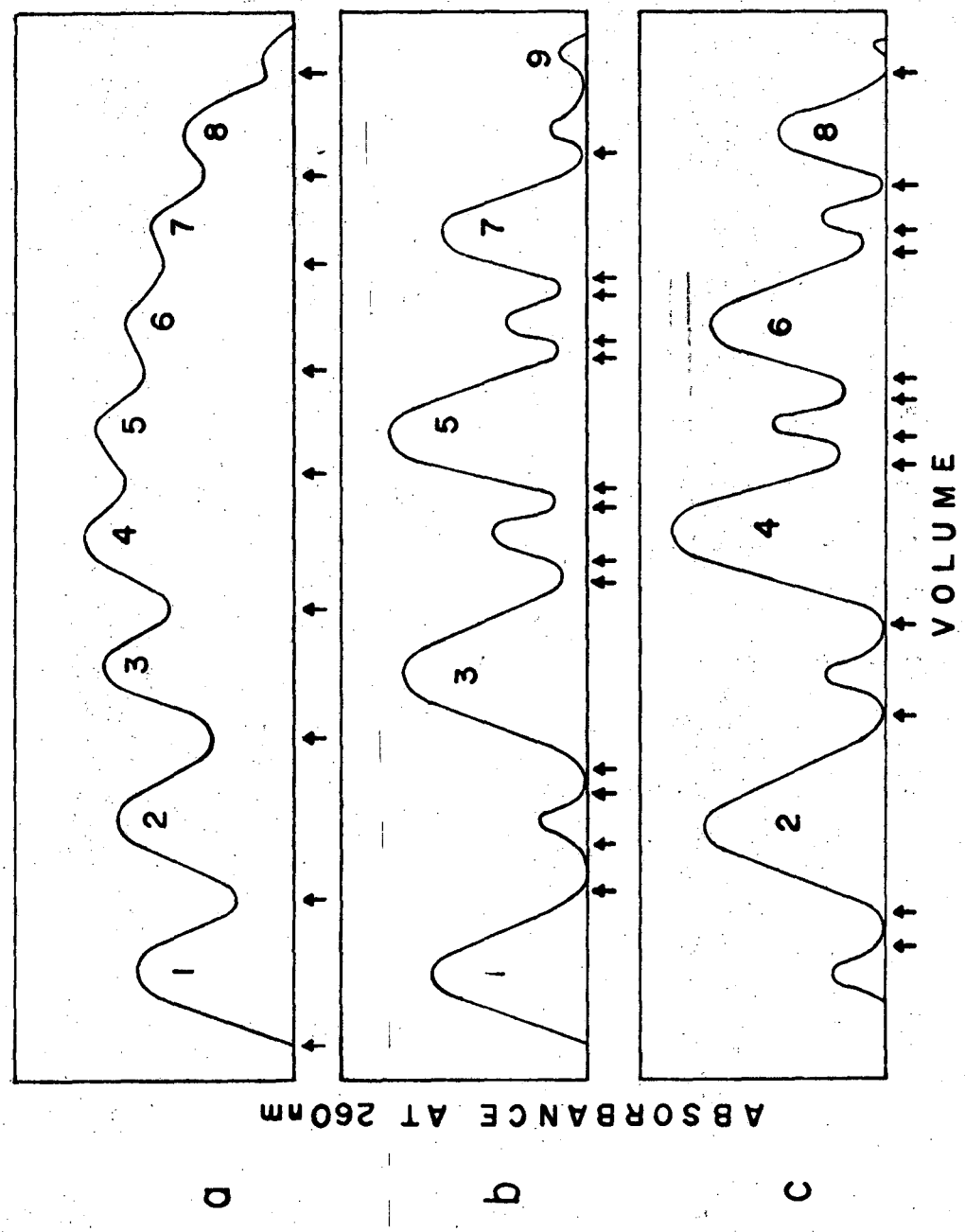


Figure 2-1

loading them all on a small (e.g. 10 to 20 ml), fat column with a high flow rate, then bombed off at a high salt concentration. The even peaks should also be concentrated in this manner. Small fast-flowing columns are also useful to concentrate individual oligomer fractions of large volume before desalting.

Column size also affects the quality of separation to some degree. This happens because an oligomer moves through the column over a range of ionic strengths rather than breaking loose at one particular salt concentration. This leads to a separation based not only on the gradient, but one rather similar to fractional distillation with the number of "theoretical plates" or ion exchange sites influencing the quality of the separation. The general rule that the sample occupy no more than the top ten percent of the column is sufficient for most purposes. For long oligomers this number should be reduced.

The flow rate is generally not an important factor in the separation when using cellulose columns. With the height of a pressure head commonly available in a laboratory rather limited, the rate is variable over only a fairly small range. A column of DEAE-cellulose packed in a 50 ml buret with a head of 2 to 3 meters will flow at 15 to 25 ml/hour. So we have generally run our columns at this flow rate. For column configurations other than the 50 ml buret (which has a gel height of about 50 cm), the column height we have used is also

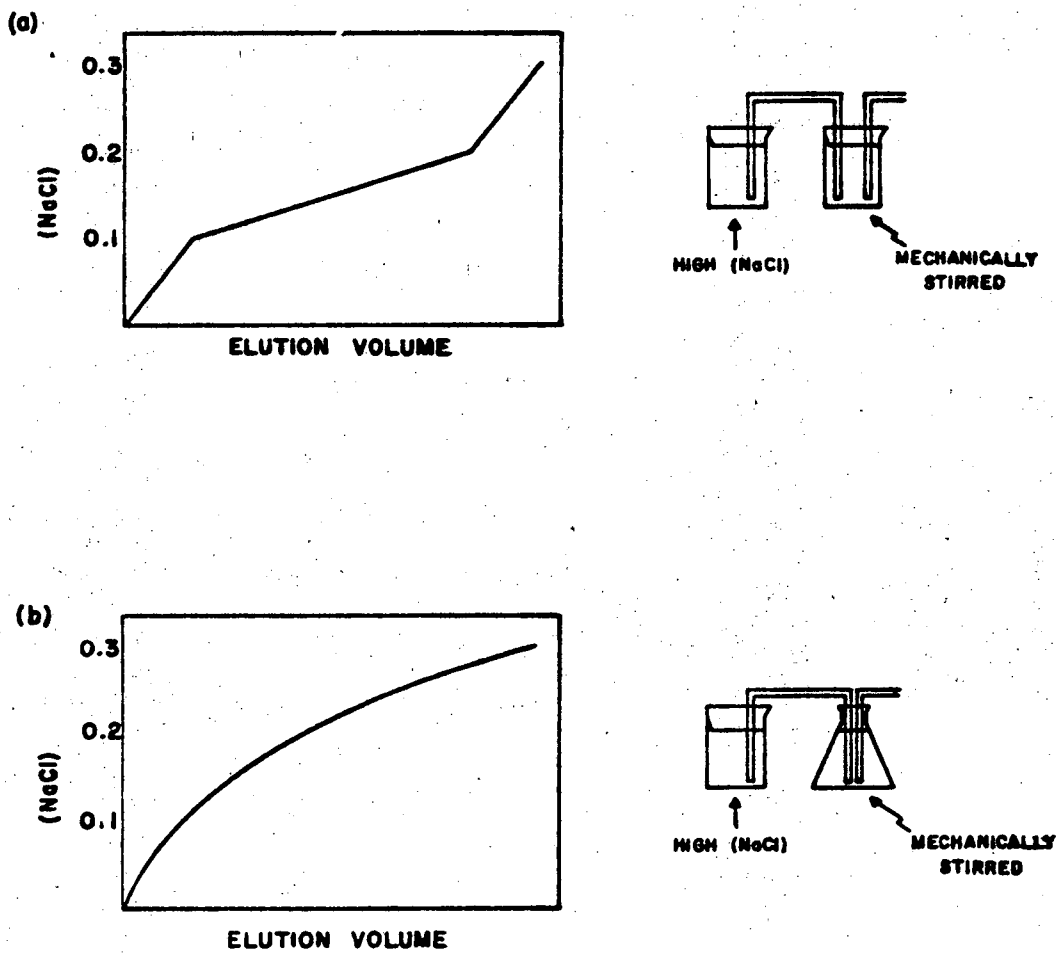
40 to 70 cm. Therefore, to a reasonable approximation, larger columns can be considered as several 50 ml columns side by side. That is, a 100 ml column should flow at about 40 ml/hour, a 25 ml column at 10 ml/hour, etc. The flow rate should be reduced when the volume between peaks will be very small since the resolution does increase somewhat with decreasing flow rate. Since the sephadex gels are beads, much higher flow rates can be achieved. The optimal flow rate still seems to be about 20 ml/hr for a 50 ml buret so the flow must be restricted.

The shape of the gradient can greatly affect the separation of a series of oligonucleotides. For the separation of oligomers with average chainlength 3 to 8 we have used a linear 0.9 liter gradient from 0 to 0.275 M NaCl, 0.01 M tris pH 8.2, 7 M urea for DEAE-cellulose and a 1.5 liter, 0.05 to 0.45 M NaCl, tris, urea gradient for DEAE-sephadex. Shallower gradients give better separations but lead to larger fraction volumes and require more purified urea. (See below.) It is frequently useful to change the slope of the gradient during the course of a separation as is illustrated in figure 2-2a. This gradient is designed to obtain a good separation of the oligomers eluting between 0.1 and 0.2 M NaCl, and skip rapidly over regions of salt concentration which are less interesting or easily separated with a steeper gradient. Another type of gradient is illustrated in figure 2-2b, where the salt concentra-

Figure 2-2.

Non-linear gradient shapes, (a) gradient skips rapidly over 0.0 to 0.1 and 0.2 to 0.3 M NaCl regions. Oligomers eluting between 0.1 and 0.2 M NaCl will be better resolved. Asymptotic gradient (b) gives better separation of long oligomers than does a linear gradient.

FIG 2-2


(a)
0.3
0.2
0.1
(NaCl)
ELUTION VOLUME
HIGH (NaCl)
MECHANICALLY STIRRED
(b)
0.3
0.2
0.1
(NaCl)
ELUTION VOLUME
HIGH (NaCl)
MECHANICALLY STIRRED

0 0 0 0 3 9 0 0 8 3

tion asymptotically approaches the high salt limit. This gradient would allow better separation of the longer oligomers than would be possible with a linear gradient, as well as keeping the fraction size of the earlier oligomers small. The asymptotic gradient is rather limited in utility since one must have carefully calibrated flasks or a gradient former in order to know the salt concentration. The salt concentration at which an oligomer elutes is a useful piece of information in the identification of the peaks. Table 2-4 lists the NaCl concentrations at which A_n (more precisely $(Ap)_{n-1}A$) and $(Ap)_n$ oligomers elute from a neutral DEAE-cellulose column in 7 M urea.[12] No matter what type of gradient is used, it should really start at about the same salt concentration as was used during packing the column since ion exchange media shrink with increasing ionic strength. If the ionic strength changes gradually the column can shrink evenly. However, a large discontinuous jump in salt concentration can cause the gel to shrink in such a manner as to form clumps and channels through it.

The pH can also be used to advantage in oligomer separation. At a pH of 3.0 to 3.5 some of the adenines (pK = 3.7 for AMP) and cytosines (pK = 4.5 for CMP) carry a positive charge to give the oligomer one less net negative charge per A or C. This has the effect of causing such oligomers to elute at a lower salt concentration. This gives a cleaner separation of oligomers,

Table 2-4. Approximate salt concentration in 7 M urea, neutral pH to elute a given Adenine homooligomer from TEAE-cellulose.*

Phosphatased oligomer	[NaCl]	Oligomer + terminal phosphate	[NaCl]
		ApAp	0.05
A_3	0.04	$(Ap)_3$	0.08
A_4	0.09	$(Ap)_4$	0.106
A_5	0.11	$(Ap)_5$	0.126
A_6	0.13	$(Ap)_6$	0.140
A_7	0.14	$(Ap)_7$	0.154
A_8	0.155	$(Ap)_8$	0.166
A_9	0.168	$(Ap)_9$	0.180
A_{10}	0.180	$(Ap)_{10}$	0.190
A_{11}	0.192	$(Ap)_{11}$	0.202
A_{12}	0.200	$(Ap)_{12}$	0.210
A_{13}	0.210	$(Ap)_{13}$	0.226
A_{14}	0.220	$(Ap)_{14}$	0.235
A_{15}	0.228		
A_{16}	0.237		
A_{17}	0.248		
A_{18}	0.255		

*Unpublished results, courtesy of Dr. O. C. Uhlenbeck.

0 0 0 0 3 9 0 0 3 8 6

since the fractional difference in charge is greater. Of course, one cannot use this technique for separations of varying length blocks of adenines or cytosines, e.g. A_6C_n and G_2A_n. We have found DEAE-cellulose to be better for fractionation at acid pH. The DEAE-sephadex has such a high exchange capacity that huge volumes of acid buffer are required to lower the pH. It is usual practice to load the sample at neutral pH, then run several volumes of acid buffer (until pH of the effluent = pH of the original buffer). Fractions should be neutralized soon after elution to prevent depurination which occurs at acid pH.

A second order effect in oligonucleotide separations is the base composition. Pyrimidine rich RNA's run slightly faster on ion-exchange media, uridine more so than cytosine. Guanine runs about the same as adenine.

Finally, we must discuss desalting in our survey of separation techniques. Most frequently we have used columns of polyacrylamide gel beads (Bio-gel P2, 50 - 100 mesh, from Bio-Rad Laboratories) to separate the oligomers from salt and urea. A 40 ml sample from the ion-exchange column is applied to the top of a 2.5 x 50 cm (ca. 250 ml) P2 column. The sample is then eluted at a flow rate of about 10 ml/min with distilled water to which 2 drops of 15 M NH_3 has been added per liter. The sample is then lyophilized, redissolved in 1 to 2 ml of distilled water and desalted again on a small P2

column (1.2 x 25 cm, ca. 30 ml). The position of the major salt fractions may be easily detected by the Schlieren effect produced when a drop of pure water is added (salt and urea solutions have a different refractive index from water) or by tasting. (Not the method of choice if cacodylate buffers have been used!) Two times desalted fractions are suitable for further PNPase additions, but for spectroscopic work where the salt concentration must be carefully controlled, it is usually the practice to desalt three times. Concentrating large fractions on small fast-flowing anion exchange columns is a useful procedure to save time. This also gets rid of the urea very quickly and thoroughly. EDTA can be removed only very poorly on desalting columns because of its high molecular weight, so it too is best removed by an initial pass through an ion exchanger at neutral pH. Trinucleotides desalt rather poorly and require many passes through P2. Dimers cannot be desalted on P2. Another desalting method is that of streaking on paper as was mentioned above. NaCl can be removed prior to developing the chromatogram in absolute ethanol.

Commercially obtained urea must be purified before use in Tener-Tomlinson type columns. The procedure we have used to remove colored material, ions, and decomposition products (CO_2 and NH_3) follows. Dissolve 25 lb. urea (B & A reagent grade, Allied Chemical) in 27 l distilled water (for 7 M urea). Mix in about 40 gm acti-

0 0 0 0 5 9 0 0 8 7

vated charcoal (Matheson, Coleman and Bell) while the urea is dissolving. Filter through an approximately 400 ml bed (in a 600 ml sintered glass funnel) of celite diatomite ("Hyflo Super-cel", Johns-Mansville) with suction from a good aspirator. If the urea preparation upon dissolving is not visibly yellow, omit the charcoal/super-cel steps. Next, run the colorless urea through a column (approximately 250 ml, 25 cm x 3.5 cm) with about 20 ml charged DEAE-cellulose or sephadex at the bottom, and filled the rest of the way with a mixed bed resin, AG 501-X8(D) 20 - 50 mesh (Bio-Rad laboratories).

D. Identification Techniques

The easiest method for identifying the oligomers produced from a primer dependent PNPase reaction is to include a small amount of radioactively labelled diphosphate in the reaction mixture. For simple one-step additions this usually means adding about 1 μc of tritiated diphosphate per μmole of cold diphosphate, and half as much for ^{14}C-diphosphates. This provides enough counts in 50 μl of each of the peaks to be 3 to 4 times background. About half as much radioactive diphosphate can be used in the nuclease assisted reactions. The 50 μl samples are first diluted with 100 μl of water to dilute the urea and then counted in 10 ml of a liquid scintillation cocktail consisting of: 1 liter toluene,

41.4 ml PPO concentrated liquid scintillator (Amersham/ Searle -- 250 gm PPO/liter toluene), and 31.6 ml BBS-3 "Bio-solv" solubilizer (Beckman). The interpretation of the counting results is absurdly easy if there is non-radioactive primer and identifiable amounts of the add one and add two peaks present. However, if the distribution is such that the first few adducts to the primer are not apparent one must resort to more sophisticated procedures. Let us take for an example the identification of the column fractions corresponding to a preparation of Pu_nPy_m, where a block of pyrimidines is added to an oligopurine primer of a specified length n. If the first visible peaks from the column UV absorption profile are obviously longer than m = 2 the peaks can be identified by their specific activities, S.A. (Actually The definition of S.A. here differs from the real specific activity by the counting efficiency. We define S.A. in terms of cpm rather than dpm.) Since S.A. equals

$$\frac{\text{cpm}}{\text{mole strand}}$$; the relative activity, R.A. equals

$$\frac{\text{cpm}}{A_{260} \cdot V_a} = \frac{\text{S.A.}}{(n+m)\varepsilon_{res}} .$$

Here V_a = the volume of the aliquot counted which has a known A_{260} and ε_{res} is the extinction coefficient of the average residue in the oligomer. If one then makes a reasonable guess as to the length of one of the counted

0 0 0 0 3 9 0 0 3 8 8

adducts to get m, $\frac{n+m}{m}$ (R.A.) ≈ constant (since S.A. is proportional to m and ε_{res} is nearly constant). One then brackets the original guess by taking a distribution one adduct shorter per peak, and then one longer. In the shorter distribution $\frac{n+m}{m}$ (R.A.) will decrease monotonically and the longer one will increase. One can also calculate the S.A. values for the distribution. The ratios of these values should then be ratios of integers.

Finally, conclusive evidence as to the identity of column fractions may be obtained by hydrolyzing some of the products. Since this is a technique destructive of the molecules which have required considerable effort to make, one usually hydrolyzes small amounts of peaks produced as by-products of the desired ones. Also, one usually gets more precise results by hydrolyzing the shorter adducts. Either hydrolysis with a specific RNase (modify the conditions given in Table 2-1 for RNase digestion to suit the oligomer in question) or complete hydrolysis in 0.3 M KOH (24 hrs at 37°C) are the commonly used procedures. With use of a specific RNase, the oligomers produced can be chromatographed on paper and their identity confirmed by comparing their mobility with known marker oligomers spotted on the same paper. In cases where the hydrolysis leads to production of identical monomers, only in varying amounts (such as the pancreatic RNase hydrolysis of A_6U_n with n greater

than 2) the ratio of the quantities of the monomers leads to the identification of the molecule. In the example above both U and Up (mixture of 2' and 3' Up) are produced, as well as A_6Up. If the U's are radioactive, the oligomers are easily characterized by their $\frac{Up}{U}$ ratio by counting the appropriate spots after paper chromatography. The identification is less certain for longer n's since the ratios are not very different from one another. The same result may be accomplished by using optical amounts of material and then eluting the appropriate spots after chromatography. This is not a very good idea for an RNaseA hydrolysis on $A_6C_5U_n$ since the Cp also produced will chromatograph very similarly to Up.

E. Treatment of Hypochromicity Data

The UV absorption maxima of base paired duplexes exhibit a characteristic temperature dependence as shown[4] in Figure 2-3a. This picture for short duplexes can be contrasted with the polymer case illustrated in Figure 2-3b for poly dAT.[23] Of course, we are most interested in quantities which give us thermodynamic parameters, therefore the evaluation of equilibrium constants and their temperature dependence are of primary importance, not absorption vs. temperature plots. One can easily convert the data in Figure 2-3b to a curve showing the

Figure 2-3.

Generating (1-f) vs. T curves from A_{260} vs. T data.

(a) A_{260} vs. T for A_5U_5[4],

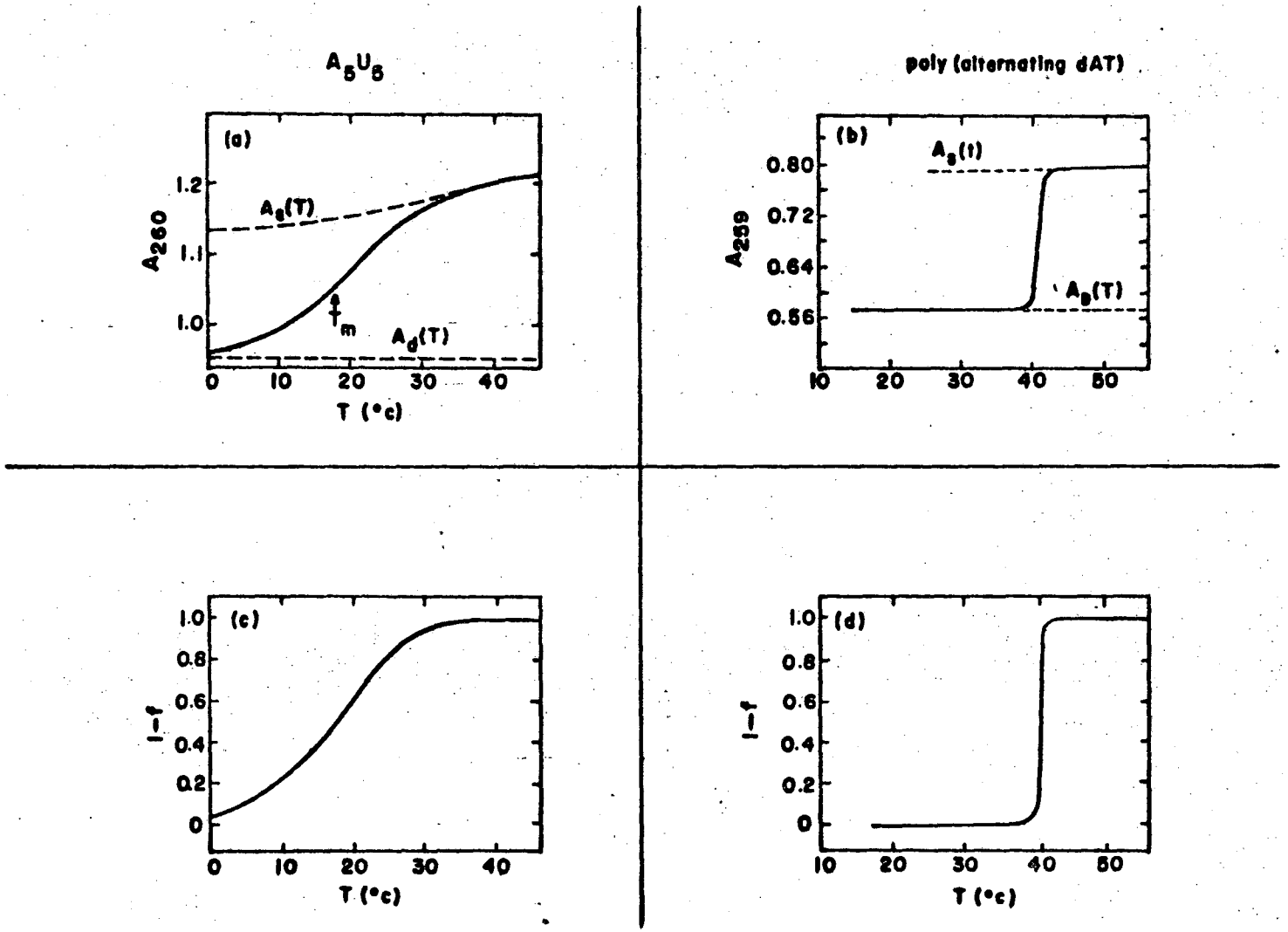
(b) A_{259} vs. T for poly dAT:AT[23],

(c) 1 - f vs. T for A_5U_5,

(d) 1 - f vs. T for poly dAT:AT.

See text for details of the conversion.

FIG 2-3


A_5U_5
(a)
A_{260}
1.2
1.1
1.0
$A_s(T)$
f_m
$A_d(T)$
0
10
20
30
40
T (°c)
poly (alternating dAT)
(b)
A_{259}
0.80
0.72
0.64
0.56
$A_s(t)$
$A_s(T)$
10
20
30
40
50
(c)
1-f
1.0
0.8
0.6
0.4
0.2
0
0
10
20
30
40
T (°c)
(d)
1-f
1.0
0.8
0.6
0.4
0.2
0
10
20
30
40
50
T (°c)

dependence of the fraction of bases paired, f, on temperature. (1-f, the fraction of single strand bases, is plotted on the ordinate since it increases from 0 to 1 with increasing temperature, rather than f which falls from 1 to 0.) The plot for poly dAT is made in Figure 2-3d. One simply takes the low temperature, zero slope portion of the curve as the absorption contribution of double strand bases, A_D; and the positive slope, upper portion of the curve as the characteristic single strand absorption, $A_S(T)$. (See Fig. 2-3c,d.) One then calculates

$$1 - f = \frac{A(T) - A_D(T)}{A_S(T) - A_D(T)} .$$

T_m is defined as the temperature at which f = 0.5, and is referred to as the melting temperature. The same analysis may be applied to the oligomeric duplex case if $A_D(T)$ is known. Frequently, however, T_m is so low that $A_D(T)$ never levels off or reaches a constant slope. In such cases one makes an extrapolation to low temperature and estimates the value of A_D. Figure 2-3c shows the 1 - f vs. T curve obtained from Figure 2-3a. The actual shape of $A_S(T)$ is somewhat more important in the oligomer case since the transition is considerably broader. The true functional dependence of A_S on T may be determined by "melting" each strand separately if that is possible, or observing the absorption-temperature

profile at a very low ionic strength where the duplex is not stable. Usually the practical difficulty in estimating $A_D(T)$ causes the largest error in T_m determinations. We believe this error to be about ± 1°C, more or less, depending on the sharpness of the transition. Also, the data can be made internally consistent by taking hypochromicity values defined by

$$\%H = \frac{A_S(40°C) - A_D(40°C)}{A_S(40°C)} \times 100$$

which are reasonable in the light of previously published data on oligomeric duplexes.[4,4a,5,6] The actual extraction of the thermodynamic parameters will be discussed in a later section but it should be noted that we now have the relative proportions of the components and can obtain the equilibrium constants.

F. Extinction Coefficient Determinations and Error Estimates

The analysis of optical data for sequence dependence relies on careful determinations of concentration. Therefore, extinction coefficients were accurately obtained by base hydrolysis or nearest neighbor calculations from dinucleoside phosphate and monomer data.[24,25] The nearest neighbor calculations were made for single strand molecules where this level of approximation is

0 0 0 0 3 9 0 0 8 9 1

sufficient to predict an accurate extinction coefficient.[26,27,28] Determination of extinction coefficients for double strand molecules were obtained by the following procedure:

1) Prepare some stock ~5 M NaOH and ~5 M HCl, 1 M Na_2HPO_4 which neutralizes an equal volume of the ~5 M NaOH to within 1%;
2) Measure A_{260} at 50°C of a sample with an A_{260} of about one;
3) Place duplicate aliquots of about 0.5 ml in preweighed 5 ml polypropylene test tubes with tight fitting caps and weigh;
4) Add ca. 30 µl of 5 M NaOH and weigh;
5) Cap tightly and wrap tube carefully with teflon tape;
6) Incubate at 34 - 37°C for about 24 hours;
7) Dry thoroughly and reweigh to check for evaporation or water pickup;
8) Add enough of the 5 M HCl / 1 M Na_2HPO_4 to exactly neutralize the NaOH (do this by calculating the weight of acid/buffer needed, adding 25 µl on the balance, and adding 1 - 3 µl increments until the addition and calculation agree within 0.3 mg);
9) Measure A_{260} of the hydrolysate at room temperature;
10) Check to see that the pH is between 6.5 and 7.5;
11) Calculate ε_{260} for the hydrolyzed mixture of monomers according to the base composition;
12) Calculate $\varepsilon_{260}^{50°}$ according to the formula

$$\varepsilon_{260}^{50°} = \frac{A\ \varepsilon_h\ V}{1.010\ A_h V_h}$$

where A = absorbance, V = volume, ε = extinction coefficient, the subscript h applies to the hydrolyzed sample, and 1.010 is a volume correction for temperature.

The following data prove useful in calculating the volume before and after the hydrolysis:

Solution	Density (25°C) ± .0005
~10 μM oligo A_6 in distilled water (OD ≈1)	1.0005
5 M NaOH	1.1780
5 M HCl, 1 M Na_2HPO_4	1.1660
1 M NaCl, 0.01 M Na_2HPO_4, 10^{-4} M EDTA (TMTB buffer)	1.0370

The densities were determined in a pycnometer of about 1 ml volume. The densities above will be only approximate for the NaOH and HCl, Na_2HPO_4, so they should be determined before accurate extinction coefficients can be determined.

We estimate the accuracy of the ε_{260} determination by the following procedure which assesses the maximum credible error. We calculate ε_{260} by the equation

$$\varepsilon = \frac{A\ \varepsilon_h\ V}{1.010\ A_h\ V_h} .$$

0 3 0 0 3 9 3 0 3 9 2

Then

$$d\ell n\varepsilon = d(\ell nA + \ell n\varepsilon_h + \ell nV - \ell nA_h - \ell nV_h)$$

or

$$\frac{d\varepsilon}{\varepsilon} = \frac{dA}{A} + \frac{d\varepsilon_h}{\varepsilon_h} + \frac{dV}{V} - \frac{dA_h}{A_h} - \frac{dV_h}{V_h} .$$

If we assume $\Delta X \simeq dX$ where ΔX is finite

$$|\Delta\varepsilon| \simeq \varepsilon \left(\left|\frac{\Delta A}{A}\right| + \left|\frac{\Delta\varepsilon_h}{\varepsilon_h}\right| + \left|\frac{\Delta V}{V}\right| + \left|\frac{\Delta A_h}{A_h}\right| + \left|\frac{\Delta V_h}{V_h}\right| \right).$$

Now if we take the reasonable values $A = 0.74$, $\varepsilon_h = 10.77$, $V = 0.5065$, $A_h = 0.747$, and $V_h = 0.5880$, we can show that $\left|\frac{\Delta V}{V}\right|$ and $\left|\frac{\Delta V_h}{V_h}\right|$ are small compared with the other terms, and

$$|\Delta\varepsilon| \simeq 9.10\left(\frac{.005}{.74} + \frac{.1}{10.77} + \frac{.005}{.747}\right) = 0.2$$

$$(|\Delta A| = |\Delta A_h| \simeq 0.005,\ |\Delta\varepsilon_h| \simeq 0.1)$$

$\varepsilon_{260} = 9.1 \pm 0.2$, about a $\pm$ 2% error.

Actually our relative errors are lower, since we always use the same set of extinction coefficients for ε_h, $|\Delta\varepsilon_h| \longrightarrow 0$. So on a relative scale our extinction coefficients should be consistent within $\pm$ 1%.

Further, we can estimate the error in a single CD measurement. The equation for determining ellipicity is $[\theta] = \frac{100\ \theta\varepsilon_\lambda}{\ell A_\lambda}$, where θ = ellipticity in degrees, ε_λ is

the molar extinction coefficient, A_λ is the absorbance at the same wavelength, and ℓ is the pathlength in cm and

$$|\Delta[\theta]| \simeq [\theta] \left(\left|\frac{\Delta\theta}{\theta}\right| + \left|\frac{\Delta\varepsilon_\lambda}{\varepsilon_\lambda}\right| + \left|\frac{\Delta A_\lambda}{A_\lambda}\right| \right) .$$

For a typical case $\theta = 0.0402$, $\varepsilon_{260} = 9.3$, $A_{260} = 0.84$, $[\theta] = 4.45 \times 10^4$, $\Delta\theta = 0.001°$ (on 0.1° full scale, wavelength ≥ 250, $A \leq 1$), $\Delta\varepsilon = \pm 0.2$, and $\Delta A = \pm 0.005$. Then

$$|\Delta[\theta]| = 4.45 \times 10^4 \left(\frac{.001}{.0402} + \frac{.2}{9.3} + \frac{.005}{.84} \right) = 0.2,$$

about ± 5%. Again, on a relative basis, the error will reduce to about ± 3%.

It should be noted that the difference between the monomer extinction coefficients in 1 M salt and distilled water at the same pH is less than 1% so the presence or absence of salt is a matter of convenience. Guschlbauer[29] reports that C deaminates to U to an extent of about 6% under these conditions. This has not been corrected for, because it usually results in a very small change. However, one should be cautioned to hydrolyze not more than 24 hours. 50°C was chosen as the point at which to get the extinction coefficients since all the molecules studied are almost completely single stranded at this temperature. This means the extinction coefficient is changing more slowly than in the melting region. The low temperature region of the melting curves is also less certain for many molecules except at high concentrations. The low temperature

0 0 0 0 3 9 0 0 8 9 3

extinction coefficient may be calculated from that at 50°C and a melting curve obtained at a concentration similar to the one at which the hydrolysis was performed. The ratio of the low temperature extrapolated absorption to the 50°C absorption times $\epsilon_{260}^{50°}$ gives the extinction coefficient of the double strand species.

G. Circular Dichroism and Ultraviolet Absorption Measurements

Circular dichroism spectra were obtained on a Cary 60 spectropolarimeter with the Cary 6001 attachment which gives the instrument CD capacity and the Cary 6003 modification which modulates the Pockels' cell at 327 Hz for a better signal/noise ratio. The light source was a xenon arc lamp with Suprasil envelope. Data was collected and processed by a small on-line digital computer as described by Cantor and Tinoco.[34] Our system used a PDP 8/s computer (Digital Equipment Corporation) and a program written by M.S. Itzkowitz and B.L. Tomlinson.[36] The spectra were digitized at 1 nm intervals from 350 to 190 nm and fit to a sliding 13 point cubic according to an algorithm by Savitzky and Golay[35] to produce smoothed curves. Further analysis was carried out on a CDC 7600 computer with the programs described in Appendix 1 at the end of this thesis. The cell compartment was modified[30] with a thermoelectric thermo-

statting device which was capable of maintaining temperatures from -15 to 100°C with an accuracy of ± 0.5°C with a stability of ± 0.04°C when 20°C tap water is used as the heat sink. Wavelength was scanned from 350 nm to 190 nm when cells of 0.1 mm pathlength were used and for spectra in distilled water in 1 cm cells, and scanned to 205 nm with 1 cm pathlengths. The noise became too high in the former case below 190 nm because of the OD due to O_2 and high oligomer absorption, and in the latter below 205 nm because of the Cl^- cutoff. The short pathlength cells had the advantage that we could work in the range of 0.01 M residue concentrations which greatly favors double strand formation. The small cells held about 6 µl, and the larger ones about 0.55 ml. Spectra were usually obtained at several temperatures in 1 M NaCl, 10^{-4} M EDTA, 0.01 M PO_4^{3-} at pH 7.0 Also, zero salt, pH 7 (adjusted with NaOH), 40°C spectra were taken. The -3.5°C, 1 M NaCl, high oligomer concentration spectra were taken as the spectra of the fully formed double strands, and the zero salt, 10^{-4} M residue concentration, 40°C spectra were taken to be the spectra of the completely single strand species. These spectra may be taken as a basis set for calculations of the type performed by Blum.[30a, 31] We used strain-free quartz cylindrical cells (Pyrocell Manufacturing Company) specially made with an outside diameter of 10 mm. These were mounted in a concentric aluminum cell holder which centered the

0 0 0 0 3 9 0 0 8 9 4

cell in the beam, reproducibly positioned the cell, and gave good thermal contact. The cells were masked with a 9 mm I.D. black rubber O-ring. Samples which were not heated above 40°C were sealed with a strip of teflon tape pressed between the stopper and ground glass stopper. The entire cell top was then wrapped in teflon tape. For samples taken to higher temperatures the cells were sealed with a stopper as shown in figure 2-4. The teflon stopper is held in place by an exterior teflon tape wrap. Before inserting the stopper, the coil of teflon tubing is filled approximately a fourth full with water at least a third of the way up the tubing (see figure 2-4). Then when the volume of the sample expands, the water in the tubing is simply forced further up the tube. The sample does not evaporate nor does it force its contents out of the space between the stopper and ground glass joint. The short pathlength cells were carefully calibrated using a basic-chromate solution standardized by the method of Cantor.[26] These cells were sealed with a small piece of Parafilm (American Can Co.) held in place over the filling slit by an approximately 5 mm length of silicone rubber tubing pressed down by a teflon plug. The cells were held in an aluminum holder described by Tinoco and Cantor.[34]

Ultraviolet absorption spectra were taken in either a Cary 14 or Cary 15 spectrophotometer. Temperature was maintained constant with a circulating bath with refrig-

Figure 2-4.

CD cell configuration for measurements above 50°C. The "water plug" moves up the teflon tubing to keep the pressure inside the cell at 1 atmosphere, minimizing evaporation.

0 0 0 0 3 9 0 0 3 9 5

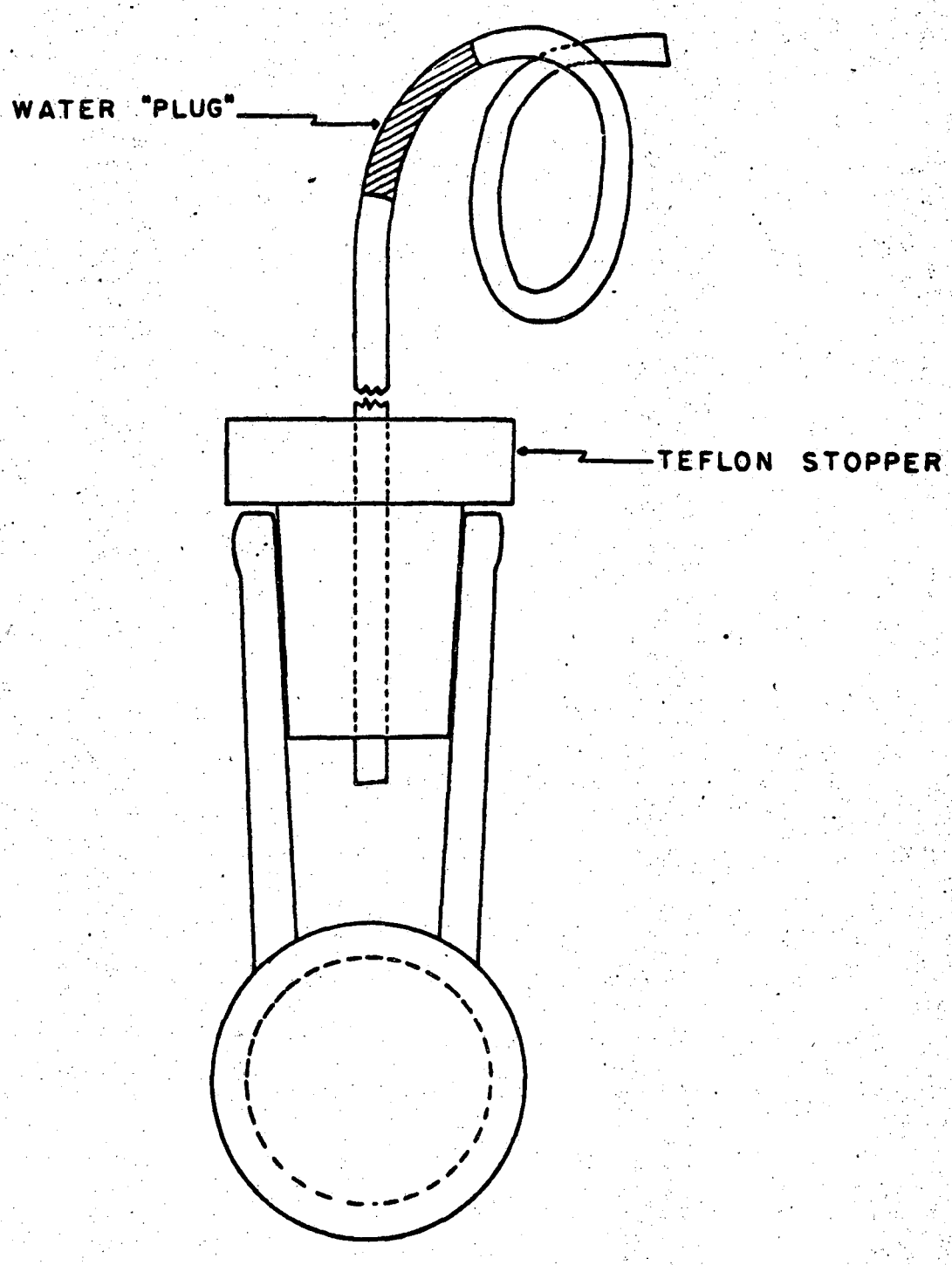


Figure 2-4

eration when necessary.

Hypochromicity data were obtained in a Gilford 2001 multiple sample recording spectrophotometer with the cell compartment temperature controlled by a Tamson circulating bath and refrigerator (Neslab Instruments). The temperature was varied by means of a small geared down electric motor which changed the setting of the bath contact thermometer at a constant rate. The temperature was increased linearly at 10 to 15 degrees per hour, and monitored by a thermocouple in the cell compartment and displayed on the chart alongside the corresponding absorption value. Some of the T_m data were hand digitized at 2° temperature intervals. The rest was obtained by running the Gilford chart strips through the Cary 60, tracing the profiles with the Cary pen, and recording the data on the PDP 8/s. The data were converted from paper tape to cards and further processed on a CDC 7600 computer.

Absorption was monitored at 260 nm. The cells (Pyrocell #1007J) varied from 1 mm to 1 cm pathlength and had thick walls so the width of the sample chamber was 4 mm. The 1 mm pathlength version held about 70 µl and the 1 cm cells had a minimum volume of about 1 ml. All cells were tightly stoppered and wrapped with teflon tape to prevent evaporation. Initially, the cells were weighed before and after each run but there was never any significant loss so the practice was discontinued.

0 0 0 0 3 9 0 0 3 9 6

Whenever it was necessary to mix two complementary strands, they were always mixed according to their carefully determined extinction coefficients at 50°C. We also did some T_m determinations on very concentrated samples (up to 20 mM in residues). For these samples we used the 1 mm cells with specially made quartz spacers (Pyrocell) 0.9 and 0.8 mm thick. These gave pathlengths of 0.1 and 0.2 mm respectively. The spaces were carefully calibrated using a standard basic chromate solution as for the short pathlength CD cells. We found it unnecessary to remove the more stubborn bubbles at the bottom of the cell since they were out of the light path and were in the same position after a run. The minimum volume for runs in such cells is 15 µl although 20 µl is considerably easier to work with. The cells were sealed with a piece of silicone rubber tubing just longer (~5 mm) than the sample slit. This was topped by a pad of teflon tape about 2 mm thick, then the stopper wedged tightly in place and held there with a tight exterior wrap of teflon tape. See Figure 2-5 for more details.

Figure 2-5.

Cell configuration for high concentration hypochromicity studies. The cell has a 1 mm pathlength and contains a 0.8 or 0.9 mm spacer. The silicone tubing is pressed tightly over the sample chamber.

0 0 0 0 3 9 0 0 3 9 7

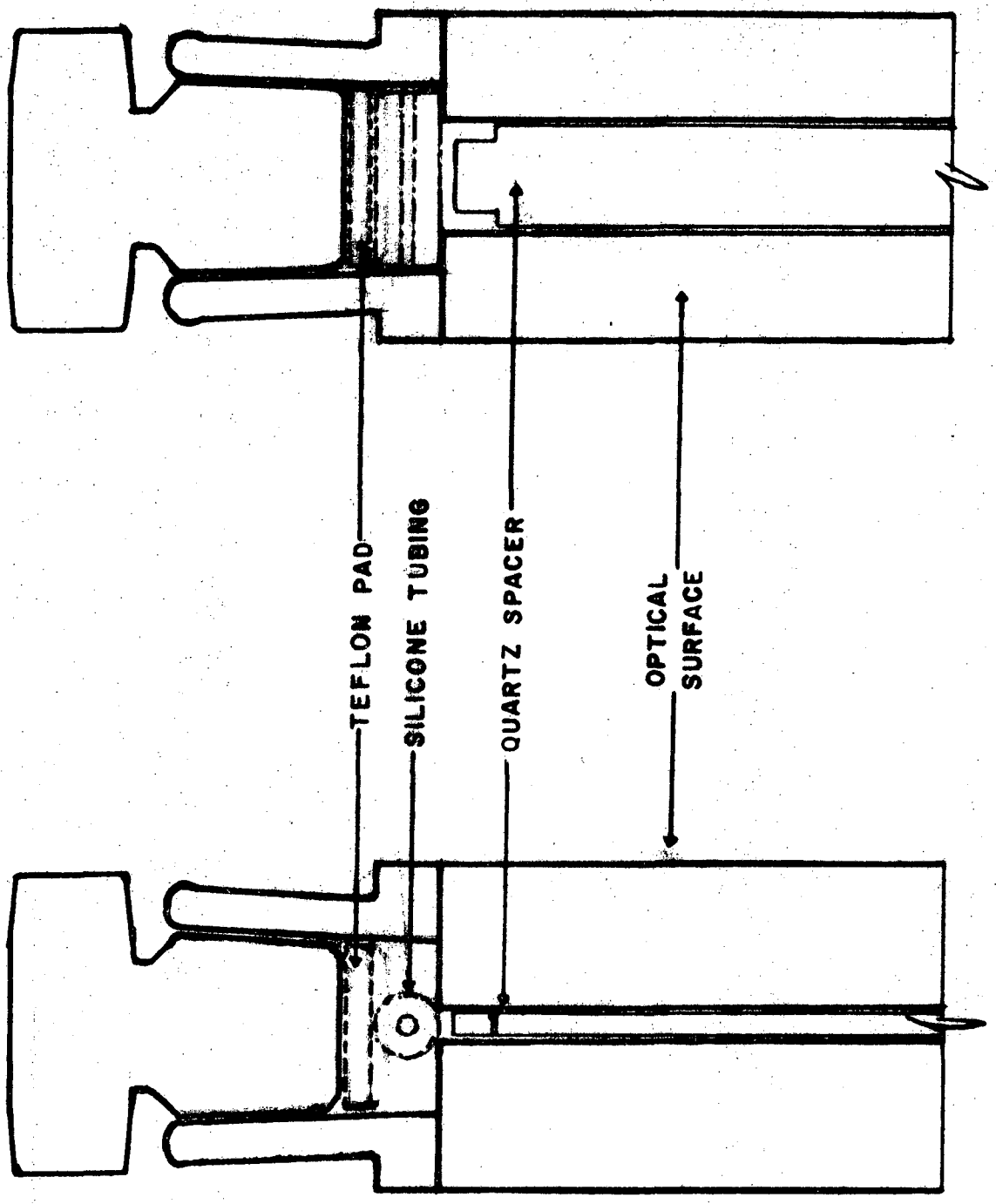


Figure 2-5

APPENDIX II - A

Sample Time Trial and Synthesis for A_4U_m

Reproduced below is a sample time trial and synthesis of A_4U_m (a simple primer dependent kinase reaction). A_4U primer was kindly supplied by Mr. Eric Wickstrom. The optimal number of A_{260} units was produced for A_4U_3 rather than the desired A_4U_4 but the desired distribution is all too seldom obtained. The trial reaction conditions were:

Volume (µl)	Substance	Stock Concentration	Final Concentration
30 *	^{3}H-UDP	0.5 mc/ml	150 µc/ml
5	UDP	0.35 M	17.5 mM
5	$MgCl_2$	0.1 M	5 mM
20	glycine-NaOH pH 9.4	1 M	0.2 M
20	NaCl	4 M	0.8 M
10	A_4U	355 A_{260} units/ml	~0.7 mM
15	PNPase	4.5 P.U./ml	0.68 P.U./ml
25	H_2O	55 M	---

* (dry down to eliminate ethanol) 100 µl total

0 0 0 0 3 9 0 0 3 9 8

15 µl aliquots were removed at 0, 30, 60, 90, 120, and 150 minutes of incubation time and spotted on Whatman 3MM chromatography paper. Also, A_4U and UDP markers were applied. Then the paper was developed over a period of about 4 days in a solvent of 6 parts 95% ethanol and 4 parts 1 M ammonium acetate. The solvent was replaced with freshly mixed solvent two times during this period as it has been observed that the composition changes over long periods of time.[12] The chromatography strip had about 40 cm of usable travel and when the UDP spots reached about 30 - 35 cm, the horizontal strip containing them was cut off and discarded. This procedure avoids radioactive contamination of the chromatography tank and allows the A_4U_n spots to travel long enough to separate. The final appearance of the chromatogram is shown in figure 2A-1. Light spots are indicated by dashed lines. The radioactive spots were then cut from the paper and immersed in 10 ml of a toluene-PPO cocktail and counted on a Beckman LS-250 liquid scintilation counter. The data are summarized in figure 2A-2. The 120 minute and 150 minute data are not included since they produced oligomers that were obviously too long. The relative A_{260} points are calculated from the total counts by taking into account the fact that a mole of A_4U_3 is twice as radioactive as the add one species, A_4U_2. Also incorporated in this calculation is the effect that more A_{260} units arise from a mole of

Figure 2A-1.

Paper chromatography of A_4U_n time trials. The UDP spots have travelled off the paper. The origin spot contains A_4U_6 and longer oligomers.

Figure 2A-2.

The cpm obtained from certain of the spots from the preceding figure are shown at the top of the figure. Relative amounts of the oligomers vs. time are shown below. Counting details and analytic procedure are given in the text.

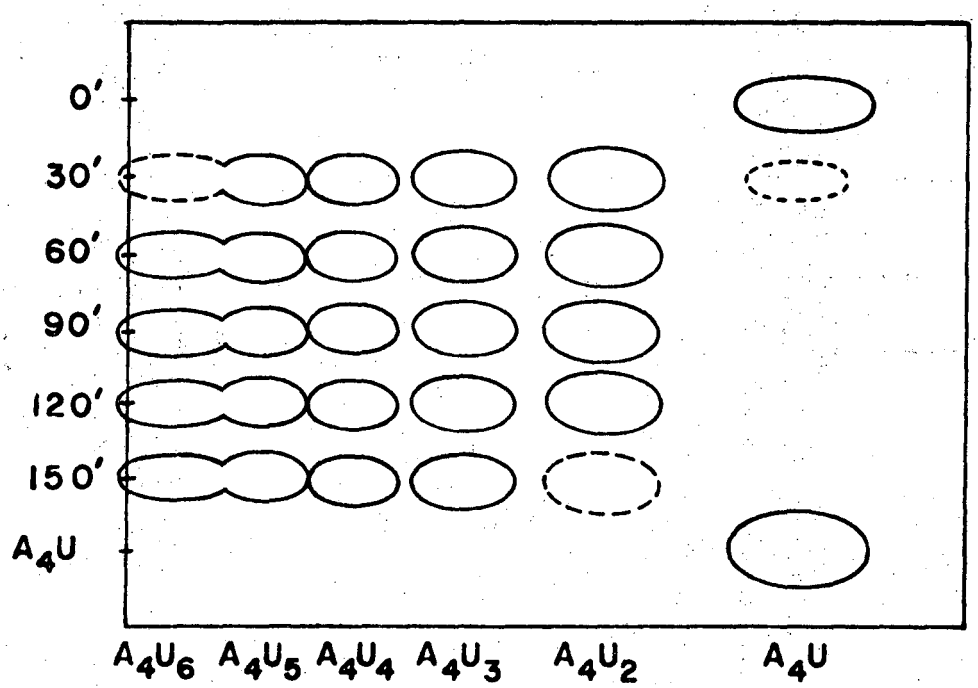
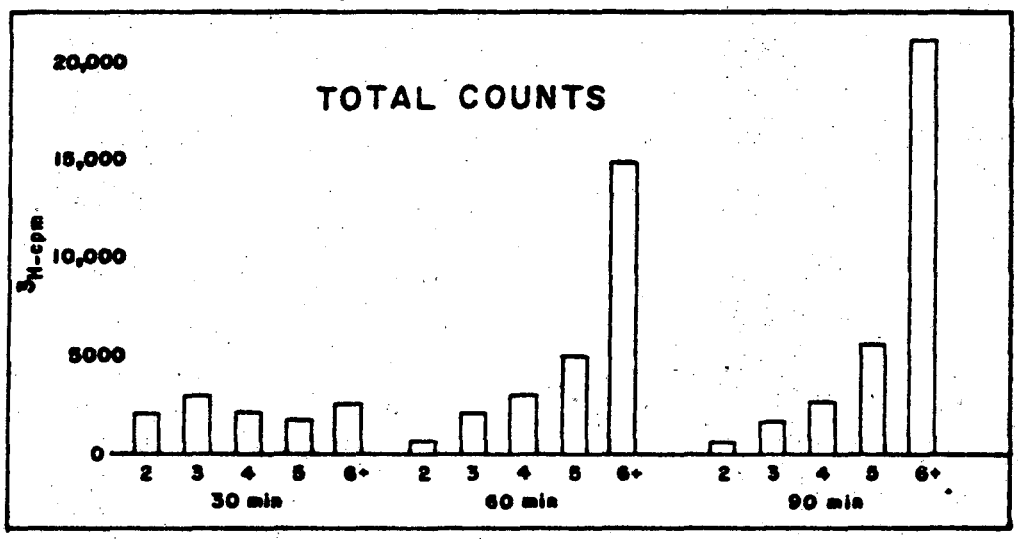
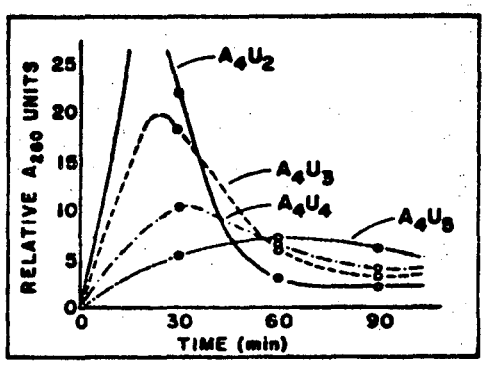


Figure 2A-1

FIG 2A-2


TOTAL COUNTS
20,000
15,000
10,000
5000
0
^{3}H-cpm
2 3 4 5 6+
30 min
60 min
90 min


RELATIVE A_{260} UNITS
25
20
15
10
5
0
0
30
60
90
TIME (min)
A_4U_2
A_4U_3
A_4U_4
A_4U_5

0 0 0 0 3 9 0 0 9 0 0

the longer species than from the shorter ones. Thus the approximate formula for the relative A_{260} units is:

$$\text{relative } A_{260} \text{ units} = \left(\frac{n+m}{m \times n}\right) \cdot \left(\frac{\text{cpm in peak m}}{\text{total cpm}}\right) ,$$

where n is the number of bases in the primer and m the length of the adduct. Such a curve for the A_4U_m trial is shown in figure 2A-2-b. The proper length of time is somewhere around the maximum in the relative A_{260} vs. time curve. It is usual practice to run the preparative reaction a somewhat longer time, however, since the resulting distributions are broader than at shorter times. Thus, if we miss the best time, usually enough of the desired peak is obtained. The time picked for the A_4U_m reaction was 45 minutes, slightly longer than the maximum in the A_{260} vs. time curve. Actually the reaction was run at 90 minutes and at half the enzyme concentration used in the trial, to conserve PNPase. The trial was scaled up 30 times to use 105 A_{260} units of A_4U primer, and the resulting column profile is given in figure 2A-3. Table 2-2 in Chapter II.B.3.e. above summarizes the reaction conditions and yields for each of the syntheses carried out in this work.

The column parameters were: 50 ml DEAE-cellulose (~50 cm × 1.1 cm), 0.9 ℓ gradient from 0 to 0.275 M NaCl, 7 M urea, 0.01 tris 8.2; flow rate ≈ 15 ml/hr.

Figure 2A-3.

DEAE-cellulose chromatography of products of the reaction A_4U + UDP $\xrightarrow{\text{PNPase}}$ A_4U_n + P_i. Reaction conditions given in text. Chromatography conditions: 50 ml (~50 cm × 1.1 cm) column, 7 M urea, 900 ml 0 to 0.275 M NaCl linear, gradient, ~5 ml fractions, ~15 ml/hr flow rate.

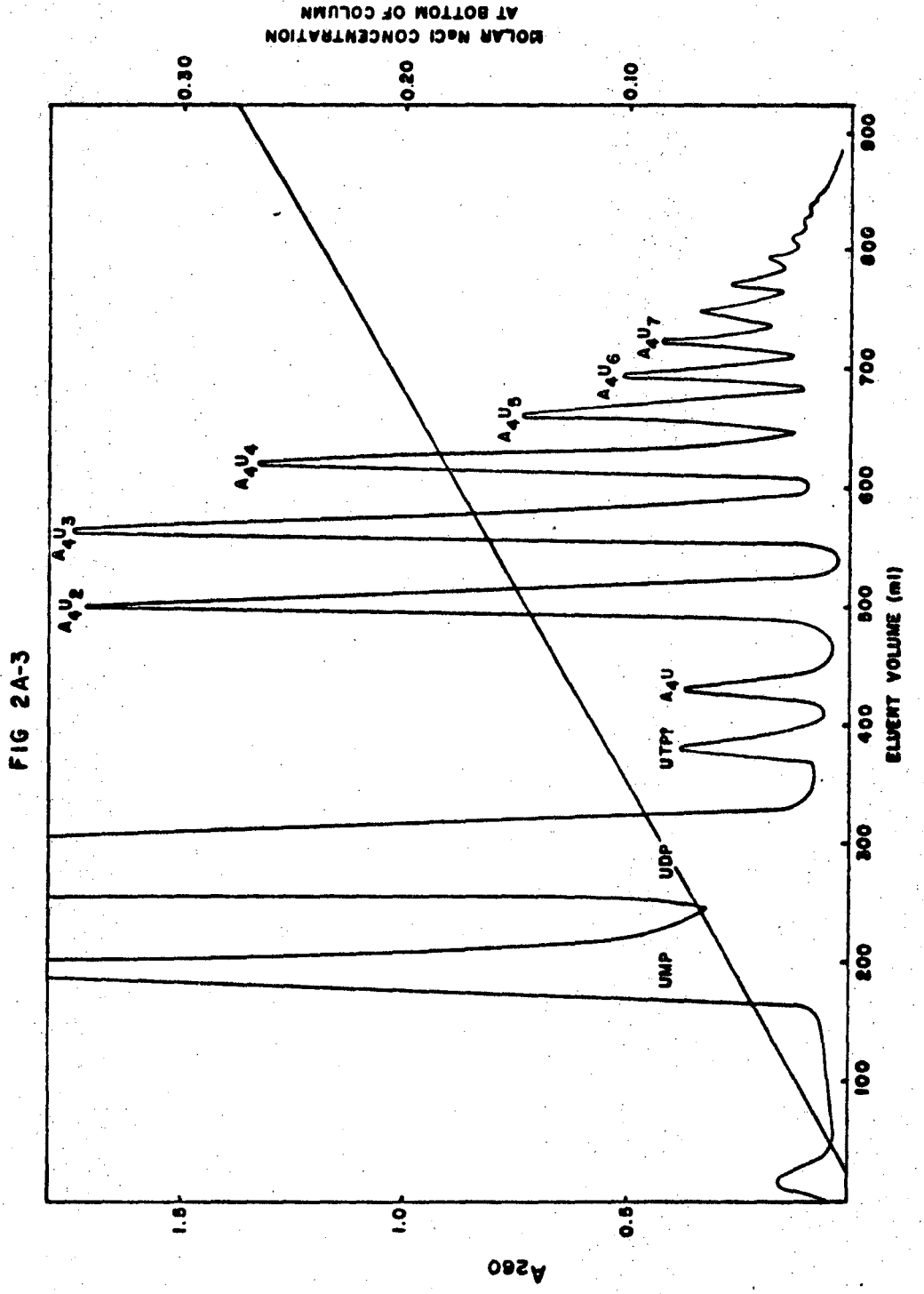

MOLAR NaCl CONCENTRATION
AT BOTTOM OF COLUMN
0.30
0.20
0.10
A260
1.5
1.0
0.5
100
200
300
400
500
600
700
800
ELUENT VOLUME (ml)
UMP
UDP
UTP?
A4U
A4U2
A4U3
A4U4
A4U5
A4U6
A4U7

FIG 2A-3

APPENDIX II - B

Primer Dependent PNPase Assay

We have spent considerable effort to try to standardize reaction conditions for the synthesis of oligo-RNA's. In studying six different preparations of primer dependent PNPase we feel that the following should comprise the standard assay.

$-A_4$	$+A_4$	
3 µl	3 µl	^{3}H-ADP (0.5 mc/ml) dried down in 3 ml conical tube...1.5µc
8 µl	8 µl	1 M glycine (pH 9.3)....0.2 M
8 µl	8 µl	4 M NaCl .....0.8 M
2 µl	2 µl	0.35 M ADP .....17.5 mM
2 µl	2 µl	0.1 M $MgCl_2$ .....5 mM
--	2 µl	oligo A_4 (A_{260} = 700)...~0.7 mM
12 µl	10 µl	H_2O
8 µl	8 µl	PNPase (1 - 2 mg/ml)....20%(v/v)
40 µl	40 µl	Total volume

Mix all of the above on ice with the exception of the enzyme. Prepare a strip of Whatman 3MM chromatography paper at least 30 cm long with a clearly marked (graphite pencil only!) origin line. Have a stream of warmed air blowing continuously over the origin line. Add the enzyme to the first tube, mix on a vortex mixer, and immediately withdraw 10 µl and spot quickly on origin line. In the same motion, start the incubation of the remainder at 37°C. Five minutes later repeat the process with the

0 0 0 0 3 9 0 0 7 0 2

second tube. Cover each tube with parafilm (American Can Company). 15 minutes after the start of each incubation, remove 10 µl more and spot. Develop the strip overnight (descending chromatography) after the spots are dried, in a solvent consisting of 30% (v/v) 1 M $NH_4C_2H_3O_2$, pH 7, and 70% ethanol (95%). Remove and thoroughly dry the paper. Cut out the areas that include A_5 and longer (usually within about 3 cm of the origin), put them in separate scintillation vials with 10 ml of a toluene/PPO cocktail and count. The counting efficiency has been carefully determined to be 4.9% for this system, and the specific activity of the ADP is 4.7×10^6 dpm/µmole. Since by definition, one unit of PNPase is the amount that incorporates 1 µmole ADP in 15 minutes, the concentration of PNPase in polymerization units per ml is: (2.17×10^{-3})(cpm at origin of 15 minute primer stimulated spot -0 minute cpm). If the enzyme is primer dependent the ratio of counts (primer stimulated: no primer) should be very high. Characteristically in this high salt concentration it is about 1000 if the enzyme is indeed primer dependent. Primer dependencies of at least 10 are required for reasonable synthesis.

We have found previous definitions of primer dependence and polymerization units to be totally unsatisfactory if one does not control the salt concentration. Also, the numbers quoted by P-L Biochemicals for their commercially available primer dependent PNPase have

very little meaning for synthesis in salt. Moreover, their assay procedure, a colorimetric assay for inorganic phosphate release has also given us a great deal of difficulty with perpetually high blank readings. In Table 2B - 1 we collect the data on polymerization characteristics of six different preparations of primer dependent PNPase.

0 0 0 0 3 9 0 0 7 0 3

Table 2B-1. Activity and Primer Dependence of several preparations of primer dependent PNPase.

	Enzyme prep[a]	Synthesis (cpm - 0 min cpm)		Enzyme concentration (P.U./ml)	Primer dependence
		+ A_4	- A_4		
no salt[b]	PNPase I	23,500	22,500	51	1.0
	PNPase II	12,000	11,900	26	1.0
	PNPase V	7,750	5,060	17	1.5
	OUCC-I	15,300	17,100	33	0.9
	PAB-2	16,300	4,500	35	3.6
0.8 M NaCl	PNPase I	2070	Back-ground	4.5	>1000
	PNPase II	610	"	1.3	>1000
	PNPase V	1560	"	3.4	>1000
	OUCC-I	890	"	1.9	>1000
	PAB-2	5090	"	11.0	>1000
	MHC	7430	"	16.1	>1000

[a]PNPase I, PNPase II, and OUCC-I were prepared by Dr. O.C. Uhlenbeck; PNPase V and PAB-2 were purchased from P-L Biochemicals and were 20 P.U./ml by P-L specifications; MHC was prepared in this laboratory by Mrs. Mei Han-Chou.

[b]PNPase V and PAB-2 had salt concentrations in the assay mixture of ∿0.02 M which may account for their slight primer dependence. The others were assayed in very low salt.

REFERENCES FOR CHAPTER II

1. Jaskunas, S.R., Cantor, C.R., and Tinoco, Jr., I., Biochemistry, 7, 3164 (1968).
2. Gennis, R.B. and Cantor, C.R., Biochemistry, 9, 4714 (1970).

2a. Pörschke, D., Biopolymers, 10, 1989 (1971).

3. Thach, R.E., in Procedures in Nucleic Acid Research, Cantoni, G.L., and Davies, D.R., eds., Harper and Row, New York (1966).
4. Martin, F.H., Uhlenbeck, O.C., and Doty, P., J. Mol. Biol., 57, 201 (1971).

4a. Uhlenbeck, O.C., Martin, F.H., and Doty, P., J. Mol. Biol., 57, 217 (1971).

5. Martin, F., Ph.D. thesis, Harvard Univ. (1969).
6. Uhlenbeck, O.C., Ph.D. thesis, Harvard Univ. (1969).
7. Mandeles, S. and Breuning, G., Biochem. Prep., 12, 111 (1968).
8. Klee, C.B., J. Biol. Chem., 242, 3579 (1967).
9. Singer, M. in Procedures in Nucleic Acid Research, Cantoni, G.L., and Davies, D.R., eds., Harper and Row, New York (1966).
10. Sachs, D.H. and Painter, E., Science, 175, 781 1972).
11. Thang, M.N., Graffe, M., and Grunberg-Manago, M., Biochim. Biophys. Acta, 108, 125 (1965).
12. O.C. Uhlenbeck, personal communication.
13. Mackey, J.K. and Gilham, P.T., Nature, 233, 551 (1971).

0 0 0 0 3 9 0 0 7 0 4

14. Tomlinson, R.V. and Tener, G.M., Biochemistry, 2, 697 (1963).

15. Lloyd, D.A., Ph.D. thesis, Univ. of Calif., Berkeley (1969).

16. Vermeulen, T. and Hiester, N.K., Ind. Eng. Chem., 44, 636 (1952).

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

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

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

20. Uchida, T. and Egami, F. in The Enzymes, P. Boyer, ed., Academic Press, New York (1971).

22. Mohr, S.C., Ph.D. thesis, Harvard Univ. (1968).

23. Inman, R.B. and Baldwin, R.L., J. Mol. Biol., 5, 172 (1962).

24. Warshaw, M.M., Ph.D. thesis, Univ. of Calif., Berkeley (1965).

25. Warshaw, M.M. and Tinoco, Jr., I., J. Mol. Biol., 20, 29 (1966).

26. Cantor, C.R., Ph.D. thesis, Univ. of Calif., Berkeley (1966).

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

28. Gray, D.M., and Tinoco, Jr., I., Biopolymers, 9, 223 (1970).

29. Guschlbauer, W., Richards, E.G., Beurling, K., Adams, A., and Fresco, J.R., Biochemistry, 4, 964 (1965).

30. Allen, F.S., Gray, D.M., Roberts, G.R., and Tinoco, Jr., I., Biopolymers, 11, 853 (1972).

30a. Blum, A.D., Ph.D. thesis, Univ. of Calif. Berkeley (1971).

31. Blum, A.D., Uhlenbeck, O.C., and Tinoco, Jr., I., Biochemistry (in press).

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

34. Tinoco, Jr., I. and Cantor, C.R., in Methods of Biochemical Analysis, vol. 18, D. Glick, ed., Wiley, New York (1970).

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

36. Tomlinson, B.L., Ph. D. thesis, Univ. of Calif., Berkeley (1969).

0 0 0 0 3 9 0 0 7 0 5

CHAPTER III

Results on Double Helices

A. The Series $\overrightarrow{A_nU_n}$ Has Only $(\overrightarrow{AA})_d$ and $(\overrightarrow{AU})_d$ Couples

These molecules form self-complementary double helices as described earlier. In a first-neighbor sense they have only two types of interactions and therefore are the simplest of molecules studied in this work. The distinct couples present are $(\overrightarrow{AA})_d$ and $(\overrightarrow{AU})_d$. Figure 3-1 presents the CD spectra obtained for A_4U_4 and A_6U_6 in 1 M NaCl, pH 7, -3.5°C and high (~0.01 M in residues) oligomer concentration. Under these conditions A_4U_4 is still approximately 10% single stranded but they all very nearly represent double strand spectra. The differences, therefore, are mainly due to the varying ratio of $(\overrightarrow{AA})_d$: $(\overrightarrow{AU})_d$. The obvious spectral characteristics are a large positive CD band at 267 nm and then several smaller positive and negative bands at lower wavelength. Both the 267 and 220 nm peaks have definitely observable shoulders. The general shape of the curves will be observed in all the double strand spectra presented later in this chapter. This is largely due to the constraints imposed by the synthetic procedure that an A_n:U_n sequence comprises a part of each of the molecules. In most it is the major sequence type present.

Figure 3-1.

Double strand CD spectra of A_4U_4 ($c_{residue}$ = 0.0104 M) and A_6U_6 ($c_{residue}$ = 0.0115 M) at -3.5°C, pH 7, [NaCl] = 1 M. The error bars indicate an estimate of the maximum relative error. (See sections I.B.4.b.ii. and II.F.)

0 0 0 0 3 9 0 0 7 0 6

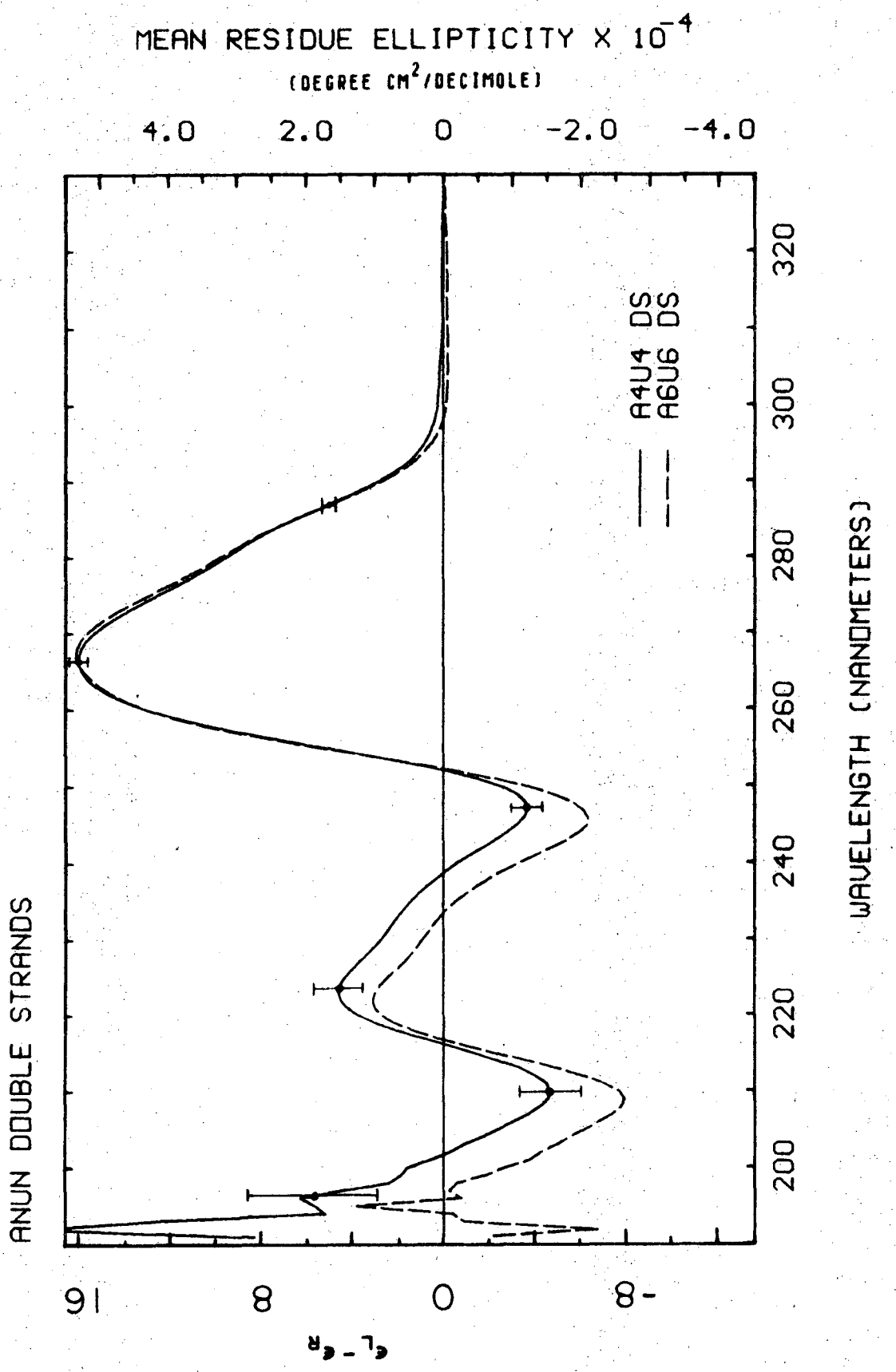


Figure 3-1

The irregular shapes of the curves below 200 nm are due to greatly increased noise at lower wavelengths. This is due mainly to the very high absorption of the samples in this region. Also the lowest 6 nm of each spectrum are not smoothed.

Figure 3-2 gives the CD spectra for the A_nU_n's under conditions where they are single stranded. A spectrum is also included for A_5U_5. Experimental difficulties with scattering prevented obtaining a high concentration spectrum. The conditions are 40°C, pH 7±0.5, $\leq$ 0.001 M NaCl, and low (~10^{-4} M in residues) concentration. We observe a considerable decrease in intensity and a red-shift from the double strand case. Again these shapes will more or less be the standard for other single strand spectra shown here.

Figure 3-3 and 3-4 compare the double and single strand spectra for A_6U_6 and A_4U_4. Also shown on these figures are spectra for the high concentration, high salt samples at several other temperatures. It is notable that the A_6U_6 changes very little with temperature, whereas the A_4U_4 changes more rapidly. Appropriate extinction coefficients are collected in Table 3-1.

B. $\overrightarrow{A_4UAU_4}$ Is Slightly More Complex

The molecule A_4UAU_4 is still self-complementary but it has a smaller percentage of the $(\overrightarrow{AA})_d$ couple than A_5U_5. In addition, it has a $(\overrightarrow{UA})_d$ type interaction

0 0 0 0 3 9 0 3 9 0 7

Figure 3-2.

Single strand CD spectra of A_nU_n's. $c_{residue} \approx 10^{-4}$, 40°C, <0.001 M NaCl, pH 7.

Figure 3-3.

CD spectra of double strand (————), single strand (·····) and partially melted (--------) A_6U_6. Double strand conditions as in Figure 3-1, single strands as in Figure 3-2. Partially melted spectra have double strand conditions except for temperature.

Figure 3-4.

CD spectra of double strand (————), single strand (·····) and partially melted (--------) A_4U_4. Double strand conditions as in Figure 3-1, single strands as in Figure 3-2. Partially melted spectra have double strand conditons except for temperature.

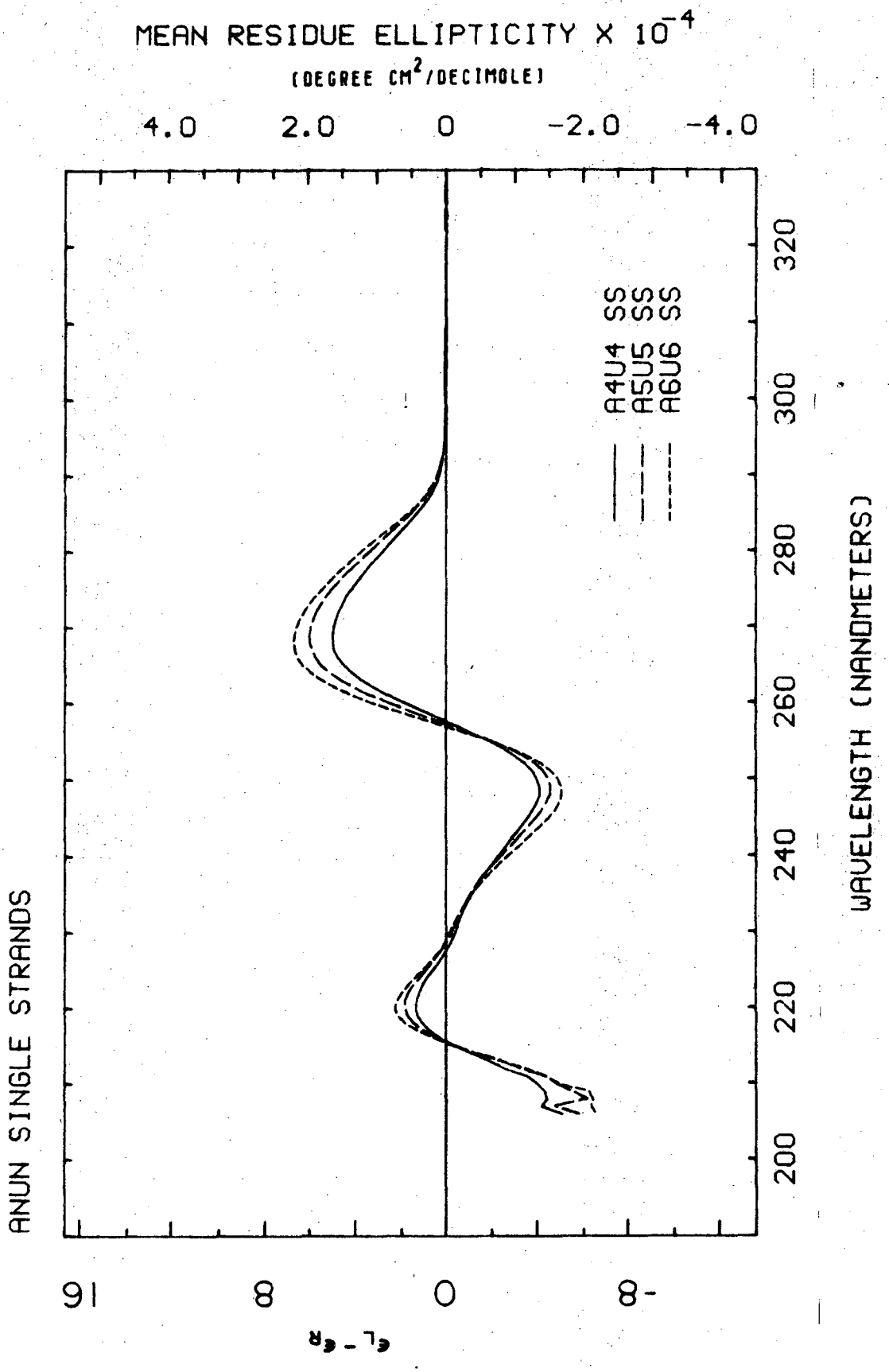

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
4.0
2.0
0
-2.0
-4.0
A4U4 SS
A5U5 SS
A6U6 SS
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
ANUN SINGLE STRANDS
16
8
0
-8
$\varepsilon_L - \varepsilon_R$

Figure 3-2

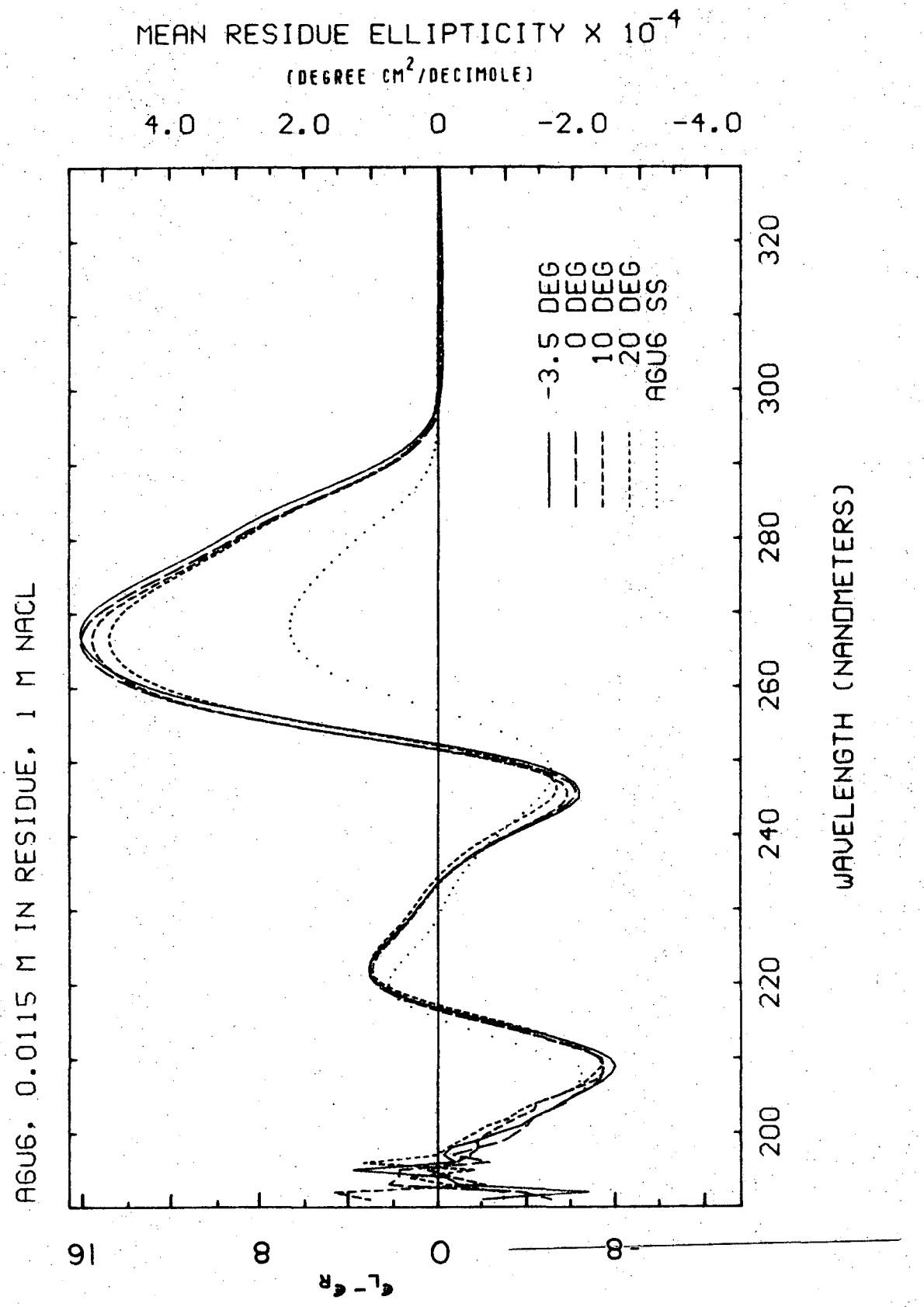

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
4.0
2.0
0
-2.0
-4.0
-3.5 DEG
0 DEG
10 DEG
20 DEG
AGU6 SS
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
AGU6, 0.0115 M IN RESIDUE, 1 M NACL
16
8
0
-8
$\epsilon_L - \epsilon_R$

Figure 3-3

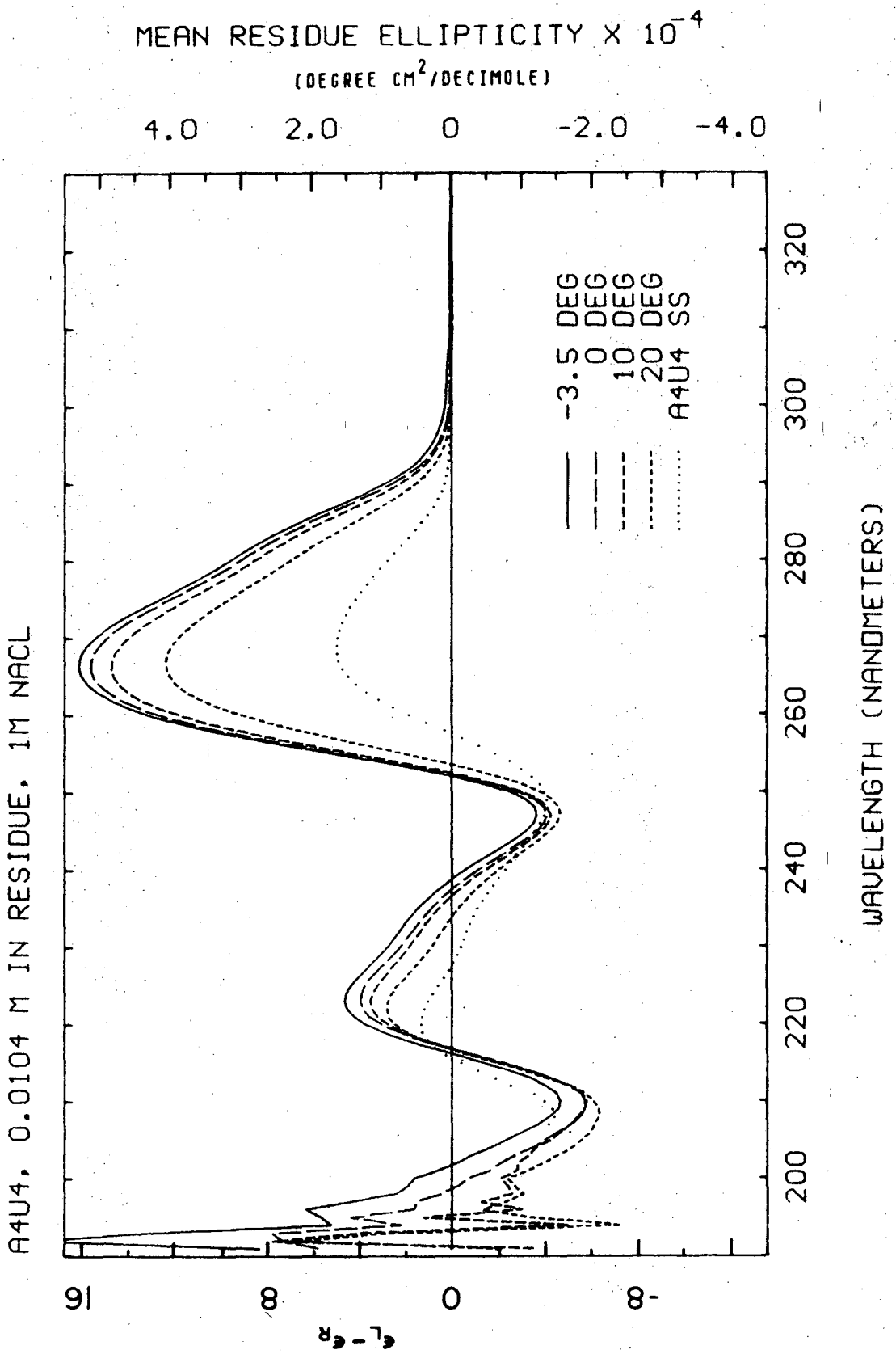

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
4.0
2.0
0
-2.0
-4.0
-3.5 DEG
0 DEG
10 DEG
20 DEG
A4U4 SS
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
A4U4, 0.0104 M IN RESIDUE, 1M NACL
16
8
0
-8
$\epsilon_L - \epsilon_R$

Figure 3-4

0 0 0 0 3 9 0 0 9 0 9

Table 3-1. Extinction coefficients for oligoribonucleotides ($x10^{-3}$)

Molecule	ε_{260} H_2O, 25°C[a]	ε_{260} 1 M NaCl, 50°C[b]	ε_{260} 1 M NaCl Double strand[c]
A_4U_4	11.14	10.26	8.03
A_5U_5	11.08	--	--
A_6U_6	11.04	10.3	7.96
A_4UAU_4	11.24	10.2	7.64
A_3CU_4	10.39	10.25	8.05
A_4GU_3	11.26		
A_5CU_4	10.71	9.99	7.79
A_4GU_5	10.95		
A_2GCU_2	10.27	10.0	8.25
A_3GCU_3	10.41	10.2	8.22
A_4GCU_4	10.50	10.2	8.28
A_5GCU_5	10.56	9.9	8.37
A_2CGU_2	10.50	9.63	8.00
A_3CGU_3	10.59	9.59	7.81
U_2CGA_2	10.53	9.9	8.42
A_3CUG	10.68	--	8.8[d]
$CAGU_3$	9.95		
A_4G_2	11.85	10.0	8.2
C_2U_4	8.72		
A_4CG	11.27	9.98	8.20
CGU_4	9.43		

[a] Calculated from dinucleoside phosphate and monomer absorptions.
[b] Measured by hydrolysis to monomers.
[c] Calculated from 50°C, 1M NaCl data, and experimental melting curve for the same molecule at the same concentration.
[d] Sample is approximately 30% single strand.

Figure 3-5.

CD of double strand ($c_{residue}$ = 0.0057, -3.5°C, pH 7, 1 M NaCl) and single strand ($c_{residue} \approx 10^{-4}$ M, 40°C, pH 7, <0.001 M NaCl) A_4UAU_4.

Figure 3-6.

CD of double strand ($c_{residue}$ = 0.0073, -3.5°C, pH 7, 1 M NaCl) and single strand ($c_{residue} \approx 10^{-4}$ M, 40°C, pH 7, <0.001 M NaCl) $A_3CU_4 + A_4GU_3$.

Figure 3-7.

CD of double strand ($c_{residue}$ = 0.0075, -3.5°C, pH 7, 1 M NaCl) and single strand ($c_{residue} \approx 10^{-4}$ M, 40°C, pH 7, <0.001 M NaCl) $A_5CU_4 + A_4GU_5$.

0 0 0 3 3 9 0 0 7 1 0

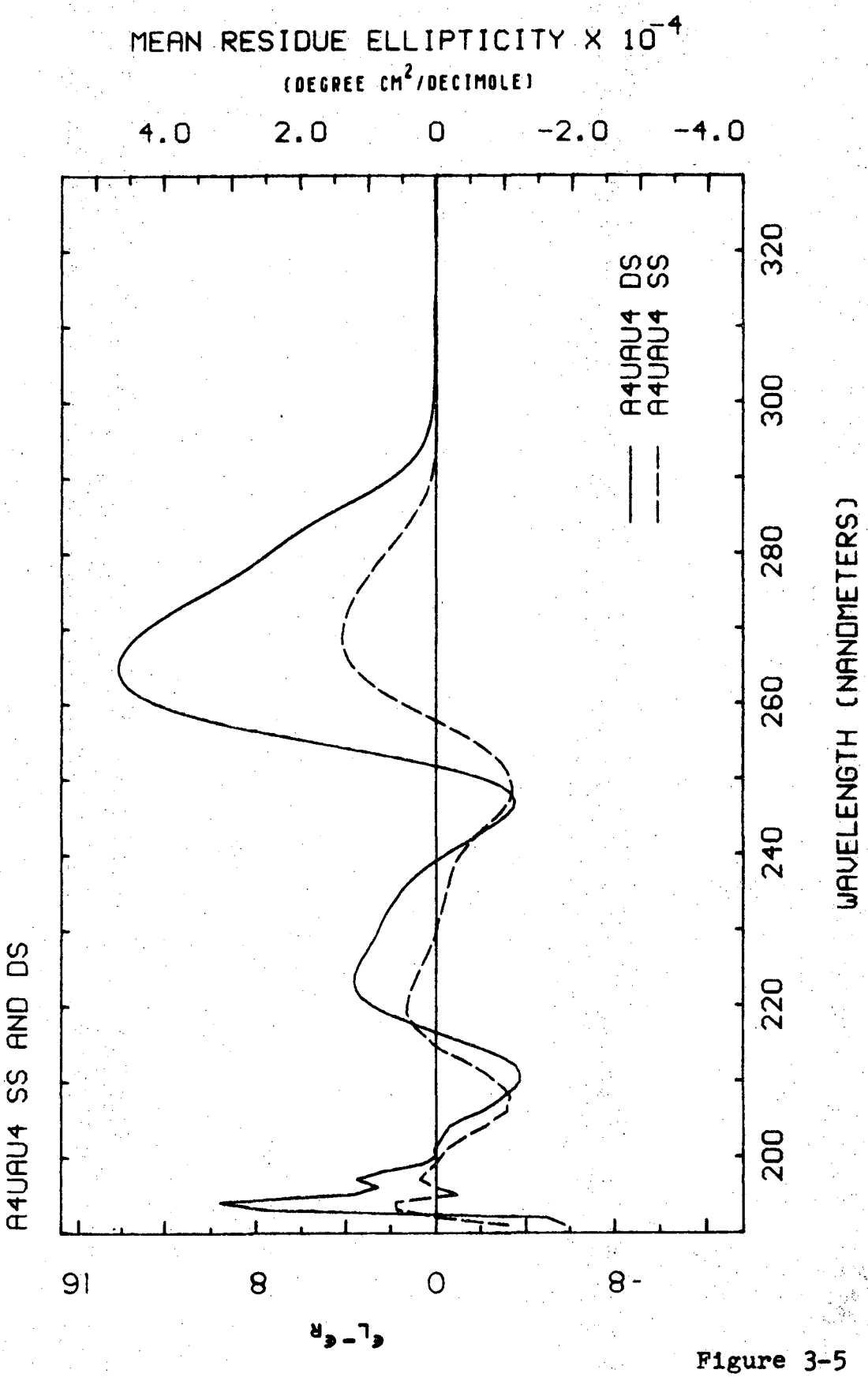


Figure 3-5

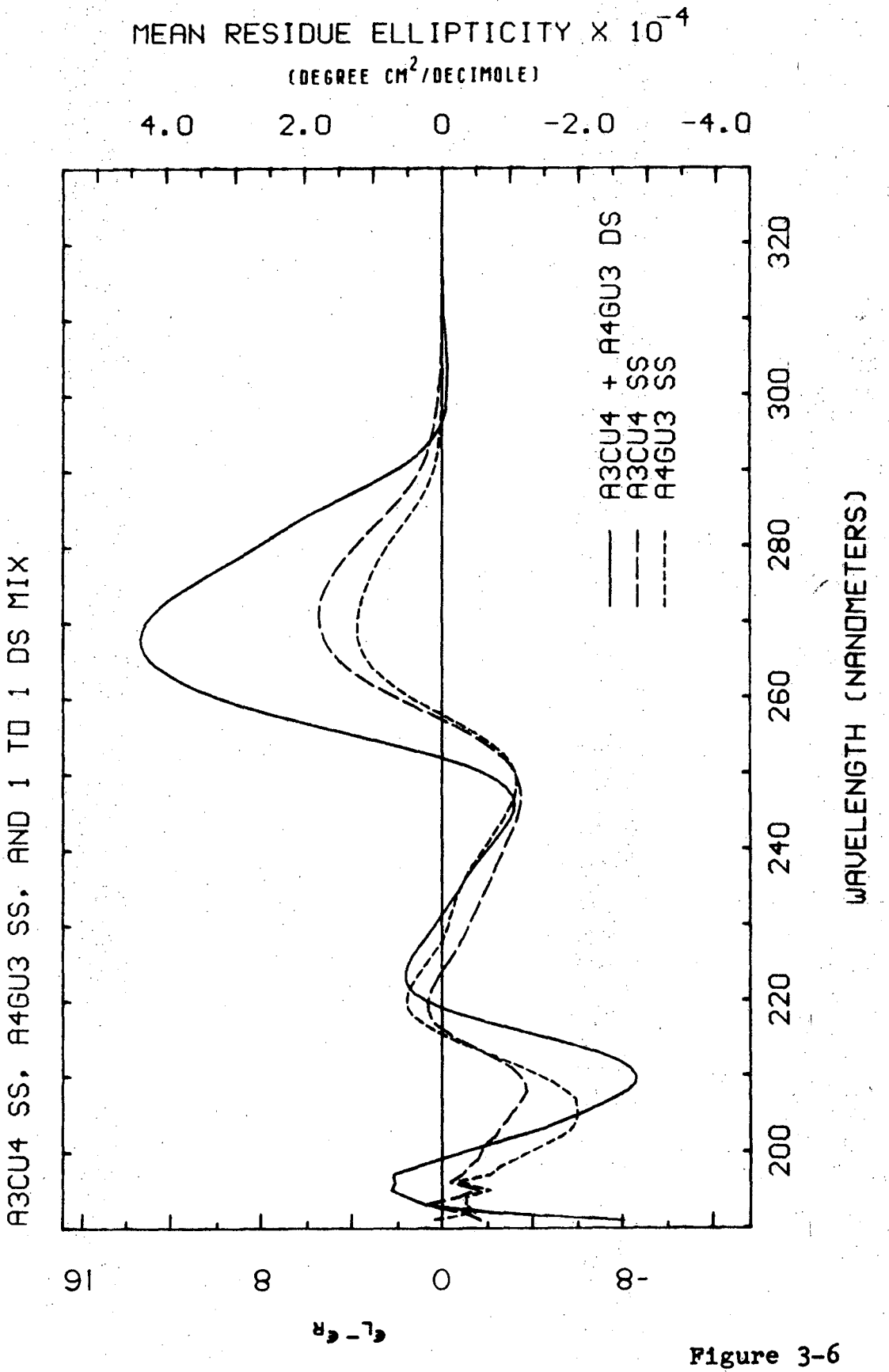

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
4.0
2.0
0
-2.0
-4.0
A3CU4 + A4GU3 DS
A3CU4 SS
A4GU3 SS
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
A3CU4 SS, A4GU3 SS, AND 1 TO 1 DS MIX
16
8
0
-8
$\epsilon_L - \epsilon_R$

Figure 3-6

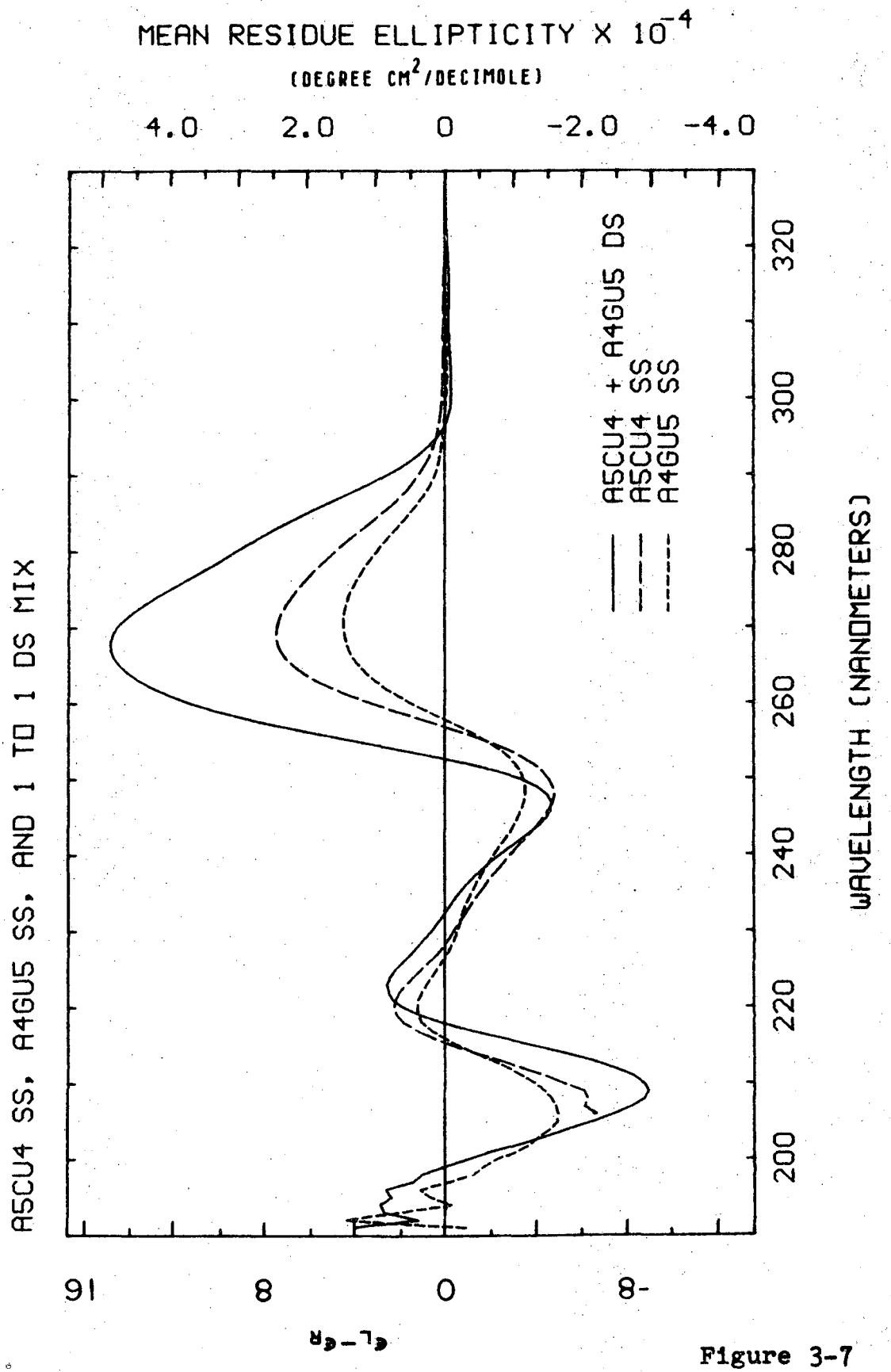

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
4.0
2.0
0
-2.0
-4.0
A5CU4 + A4GU5 DS
A5CU4 SS
A4GU5 SS
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
A5CU4 SS, A4GU5 SS, AND 1 TO 1 DS MIX
16
8
0
-8
$\epsilon_L - \epsilon_R$

Figure 3-7

present. Figure 3-5 gives the double and single strand spectra for this molecule. The double strand has a much more pronounced shoulder on the 220 nm peak than observed in the A_nU_n's and the long wavelength maximum is shifted about 2 nm to the blue. In neither of these first two series do we observe a band centered around 300 nm. The spectra at high concentration do not change rapidly with temperature below 20°C. This is also true of the data to follow unless indicated otherwise.

C. $\overrightarrow{A_nCU_m} + \overrightarrow{A_mGU_n}$ Have One GC Base Pair

In this series we have studied the complementary molecules $A_3CU_4 + A_4GU_3$ and $A_5CU_4 + A_4GU_5$. The former still has five $(\overrightarrow{AA})_d$ couples out of its seven first-neighbor interactions so both spectra still are quite similar to the A_nU_n's. (See Figures 3-6 and 3-7.) The most striking effect of introducing the G-C base pair is to produce a very small, but real negative band at about 300 nm. Also included in the figures are the single strand spectra.

D. $\overrightarrow{A_nGCU_n}$ Have Two G-C Base Pairs

This series is represented here by four molecules with n = 2,3,4 and 5. These and the A_nCGU_n series are the most easily synthesized molecules with 3 distinct first-neighbor couples. Also, having two G-C base pairs, even the short members are quite stable. The double

0 0 0 0 3 9 0 0 7 1 2

strand spectra are presented in Figure 3-8. They are quite similar to the A_nU_n spectra with some striking exceptions. A trough appears at about 300 nm, whose intensity increases with the GC content. The 245 nm trough decreases in magnitude, and broadens with increasing GC content. The 210 nm trough is considerably deeper for this series than for the A_nU_n's. The trends in the 267 and 220 nm peaks are not apparent, although as n increases we expect them to approach the A_nU_n values. It is our feeling, based on experiments previously performed at lower concentrations in 1 M NaCl that the 267 nm peak is rather insensitive to GC content for this sequence, and the 220 nm band drops. It is likely then that the n = 3 and 4 members of the series are somewhat in error. The situation is also complicated by the fact that the experimentally determined extinction coefficients (table 3-1) do not decrease with n as expected. It is not surprising that A_2GCU_2 shows substantial spectral differences from A_nU_n since it has only 40% of its first-neighbor couples as $(\overrightarrow{AA})_d$. The single strand spectra are given in Figure 3-9.

E. $\overrightarrow{A_nCGU_n}$ Have Two G-C Base Pairs But Different First Neighbor Couples

The double strand spectra for A_2CGU_2 and A_3CGU_3 are given in Figure 3-10. They are also considerably different from the A_nU_n data. Focussing on A_2CGU_2 we

Figure 3-8.

CD of double strand A_nGCU_n's ($c_{residue}$ = 0.0089, 0.0079, 0.0092, and 0.0140 for n = 2,3,4,5 respectively; -3.5°C, pH 7, 1 M NaCl).

Figure 3-9.

CD of single strand A_nGCU_n's ($c_{residue}$ $\approx 10^{-4}$ M, 40°C, pH 7, <0.001 M NaCl).

Figure 3-10.

CD of A_nCGU_n double strands ($c_{residue}$ = 0.0090 and 0.0066 for n = 2 and 3 respectively, -3.5°C, pH 7, 1 M NaCl), and single strands ($c_{residue}$ $\approx 10^{-4}$ M, 40°C, pH 7, <0.001 M NaCl).

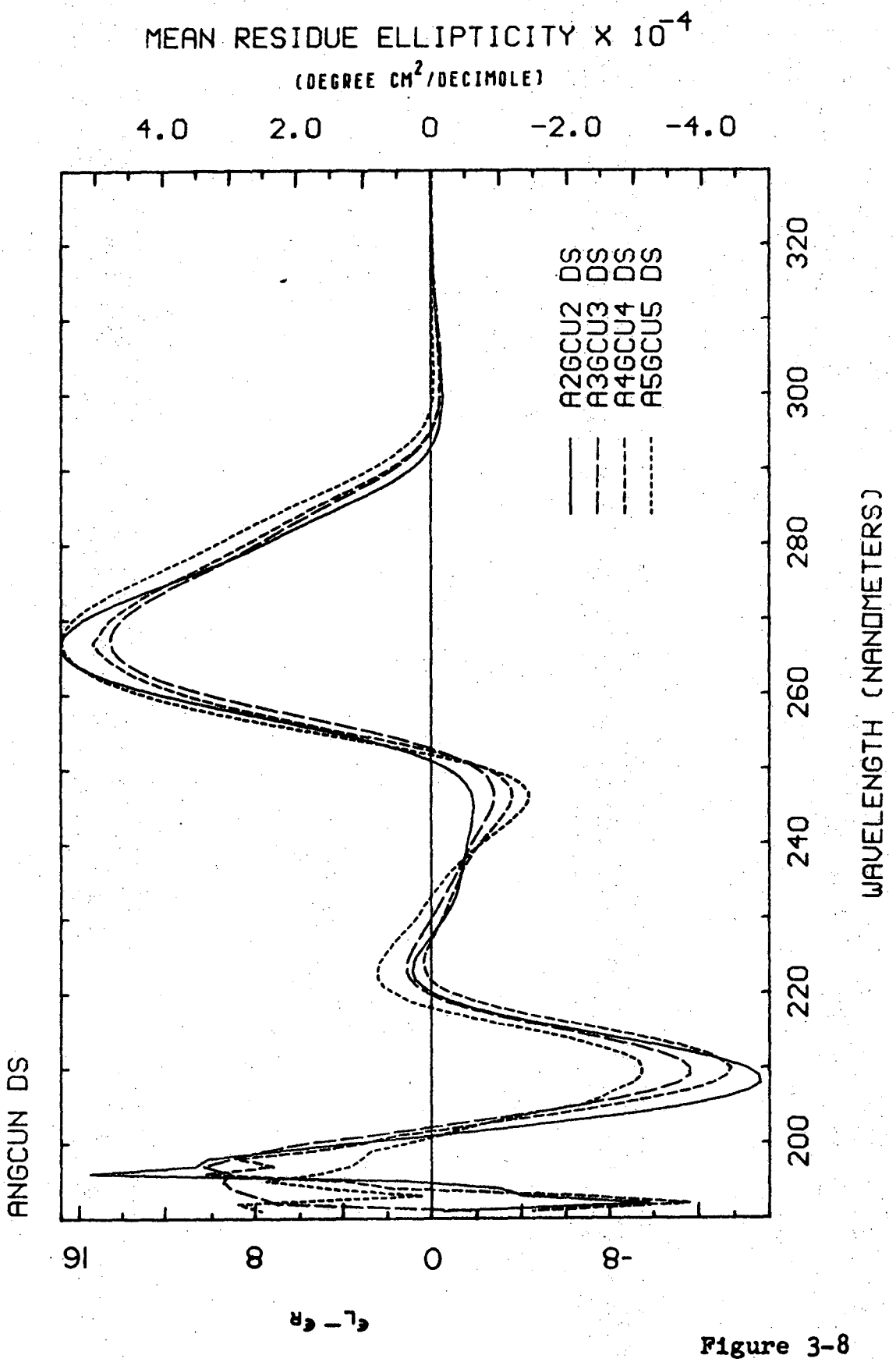

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
4.0
2.0
0
-2.0
-4.0
A2GCU2 DS
A3GCU3 DS
A4GCU4 DS
A5GCU5 DS
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
ANGCUN DS
16
8
0
-8
$\epsilon_L - \epsilon_R$

Figure 3-8

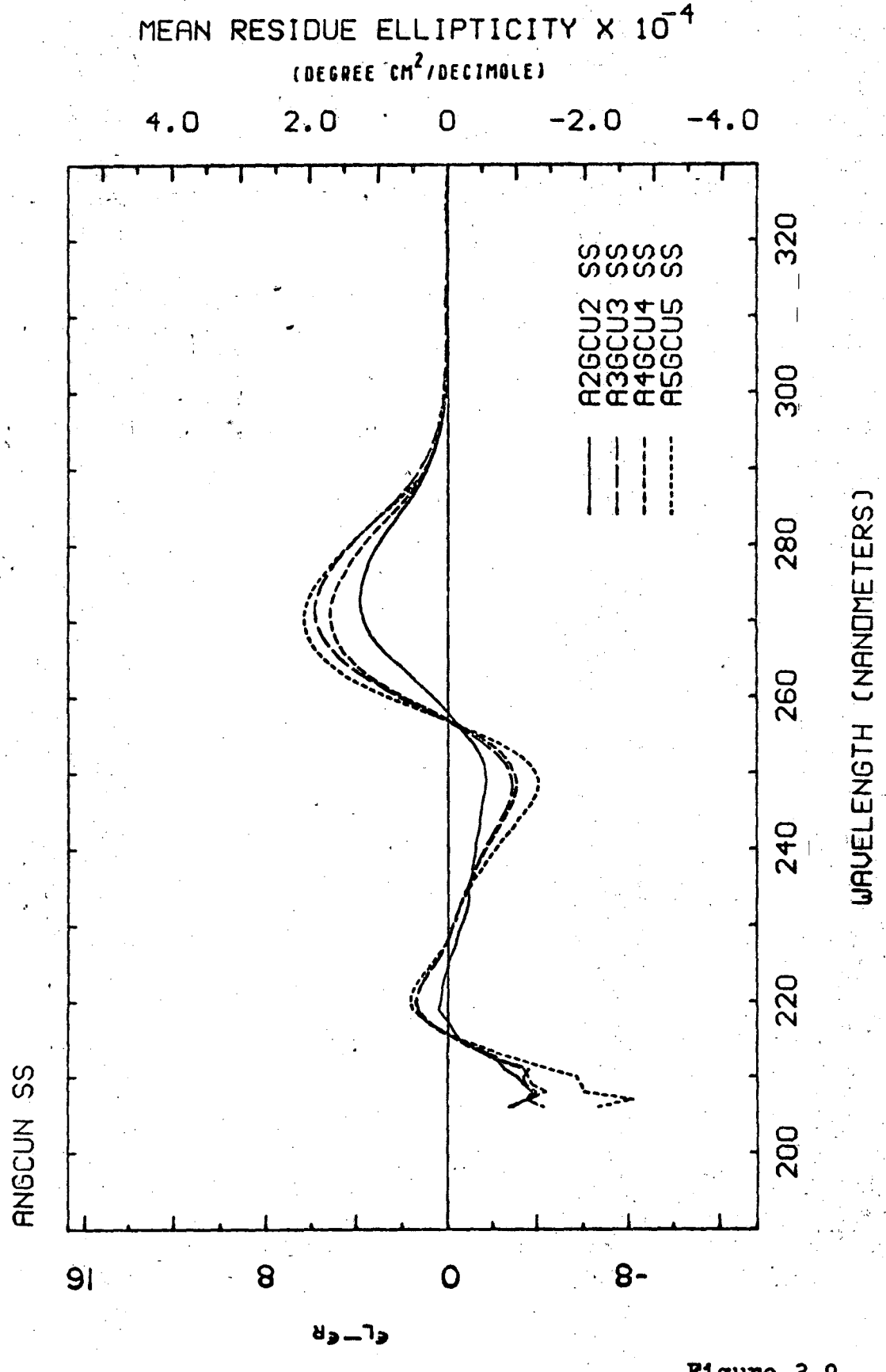

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
4.0
2.0
0
-2.0
-4.0
A2GCU2 SS
A3GCU3 SS
A4GCU4 SS
A5GCU5 SS
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
ANGCUN SS
16
8
0
8-
$\epsilon_L - \epsilon_R$

Figure 3-9

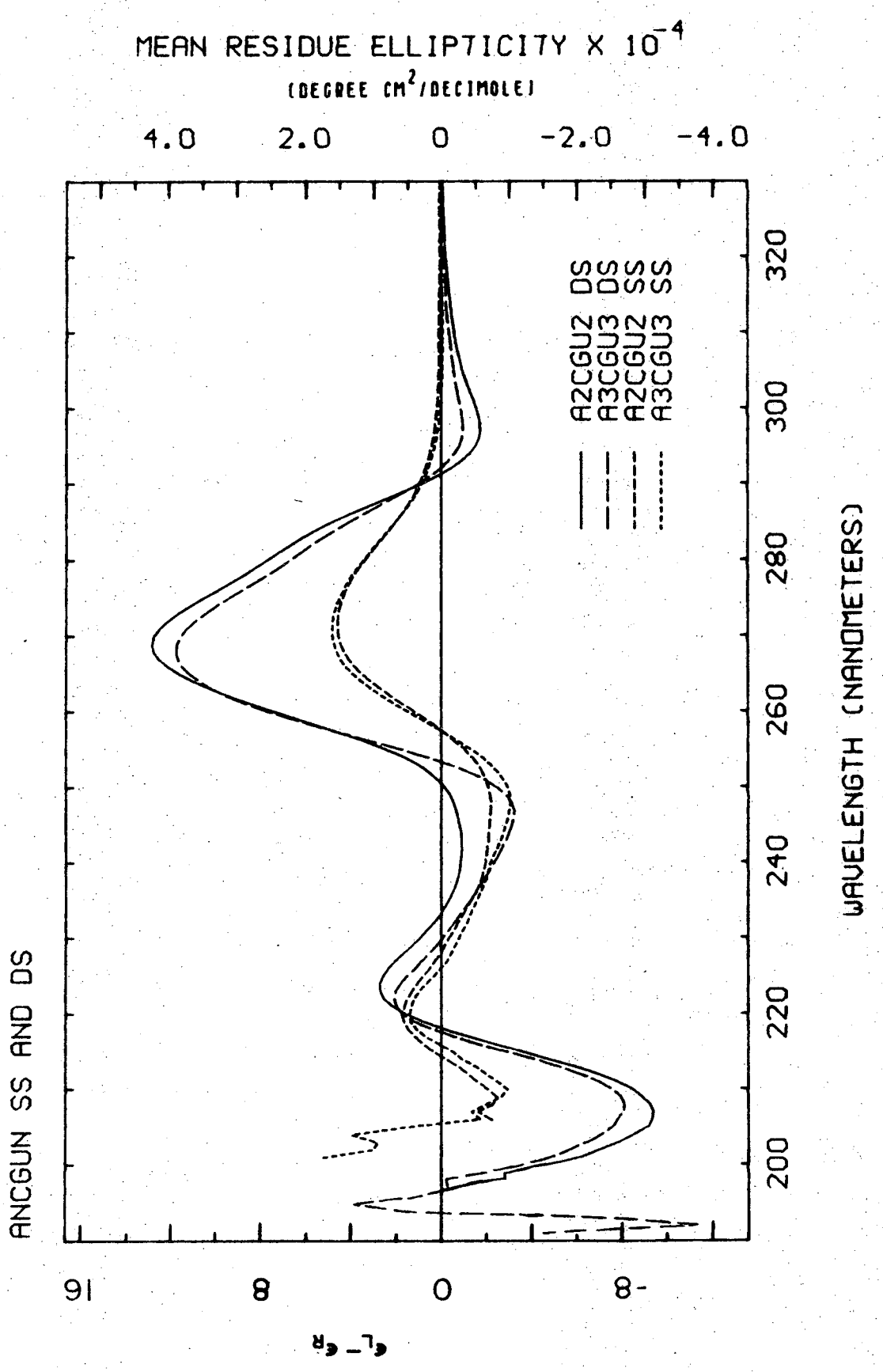

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
4.0
2.0
0
-2.0
-4.0
A2CGU2 DS
A3CGU3 DS
A2CGU2 SS
A3CGU3 SS
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
ANCGUN SS AND DS
16
8
0
-8
$\epsilon_L - \epsilon_R$

Figure 3-10

see the largest trough at 300 nm of all the double helices we have obtained. The 267 nm magnitude is lower and a very broad trough at 245 nm is seen. The 210 nm trough is quite a bit broader and deeper than A_nU_n. Diluting out the $(\overrightarrow{AC})_d$ and $(\overrightarrow{CG})_d$ couples with more $(\overrightarrow{AA})_d$ as in A_3CGU_3 brings a closer agreement with A_nU_n as expected. The single strand spectra are also given in Figure 3-10.

F. $\overrightarrow{U_2CGA_2}$ Is Also A Sequence Isomer Of A_2GCU_2

Figure 3-11 gives the U_2CGA_2 double strand and single strand spectra. The arrangement of chromophores in this molecule leads to a remarkable low intensity for the long wavelength maximum. There is no long wavelength negative band, the ~245 nm trough is shallow, and the 210 nm trough is quite broad. It is possible that the molecule is not totally in the duplex form. However, there is good evidence to the contrary. First, there is a 4 nm shift to the red in the crossover around 255 nm in going from the double to single strands. This compares reasonably to the 5 - 6 nm shift seen in the A_nU_n's. Also, the temperature dependence of hypochromicity (unpublished) shows the molecule to have a T_m of about 25°C at this concentration with a simple monophasic melting curve and hypochromicity of about 15%. This compares with about 17 - 18% hypochromicity for other double helices with the same base composition.

0 0 0 0 3 9 0 0 7 1 5

Figure 3-11.

CD of U_2CGA_2 double strand ($c_{residue}$ = 0.0090 M, -3.5°C, pH 7, 1 M NaCl) and single strand ($c_{residue}$ $\approx 10^{-4}$ M, 40°C, pH 7, <0.001 M NaCl).

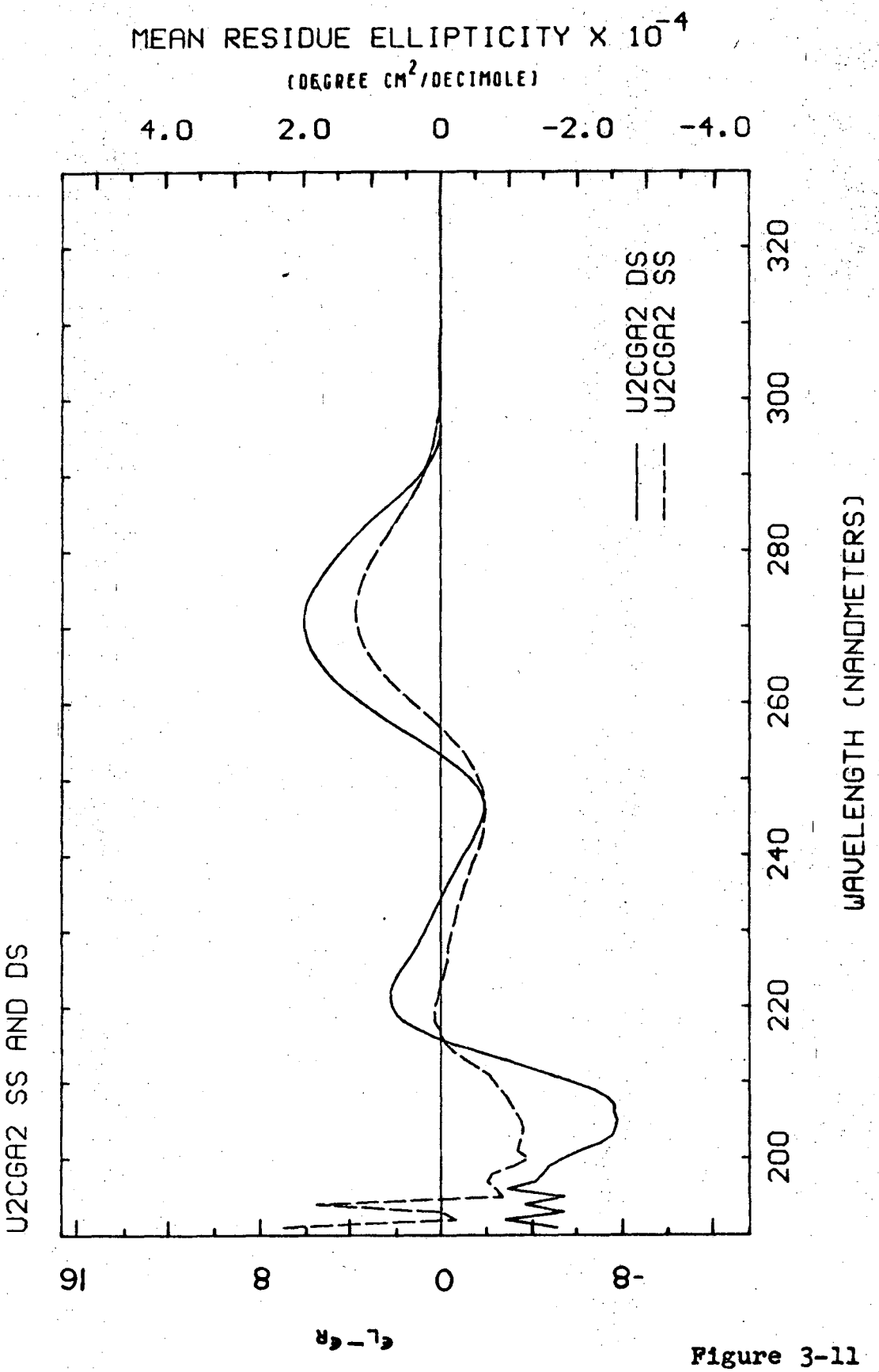

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
4.0
2.0
0
-2.0
-4.0
U2CGA2 DS
U2CGA2 SS
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
U2CGA2 SS AND DS
16
8
0
-8
$\epsilon_L - \epsilon_R$

Figure 3-11

0 0 0 0 3 9 0 0 7 1 6

G. $\overrightarrow{A_3CUG}$ + $\overrightarrow{CAGU_3}$ Have Two Non-adjacent G-C Base Pairs

This arrangement of base pairs unfortunately leads to a relatively unstable double helix and at the concentrations used here is about 30% single strand. Even more unfortunate is the fact that so far we have no other way of getting the CD contribution, $[\overrightarrow{CA}]_d$. We have decided to proceed with this and revise as necessary when further experiments are completed. Figure 3-12 shows the ~70% DS spectrum along with its constituent single strands at 40°C, 1 M NaCl. It seems likely that there is at least a small negative band at about 300 nm. Further spectral details are difficult to discern because of the degree of helix melting. Note the differences in the single strand spectra A_3CUG and $CAGU_3$ in Figure 3-12.

H. $\overrightarrow{A_4G_2}$ + $\overrightarrow{C_2U_4}$ Have A G-C Base Pair On The End

This double helix is quite stable and one does not encounter the problems mentioned above. The high concentration spectra are reasonably constant from -3.5°C to 10°C. This is the only molecule with the couple $(\overrightarrow{CC})_d$ in addition to $(\overrightarrow{AA})_d$ and $(\overrightarrow{AG})_d$. The double and single strand spectra are shown in Figure 3-13. The duplex CD has a definite but very small trough at about 300 nm, which has been consistently observed. The other spectral characteristics are similar to A_2CGU_2 with a broad 210 nm band. The single strand spectra show

Figure 3-12.

CD of A_3CUG + $CAGU_3$ ~70% double strand ($c_{residue}$ = 0.0131 M, -3.5°C, pH 7, 1 M NaCl) and single strands ($c_{residue} \approx 10^{-4}$ M, 40°C, pH 7, <0.001 M NaCl).

Figure 3-13.

CD of A_4G_2 + C_2U_4 double strand ($c_{residue}$ = 0.0079 M, -3.5°C, pH7, 1 M NaCl) and single strands ($c_{residue} \approx 10^{-4}$ M, 40°C, pH 7, <0.001 M NaCl).

0 0 0 0 3 9 0 0 9 1 7

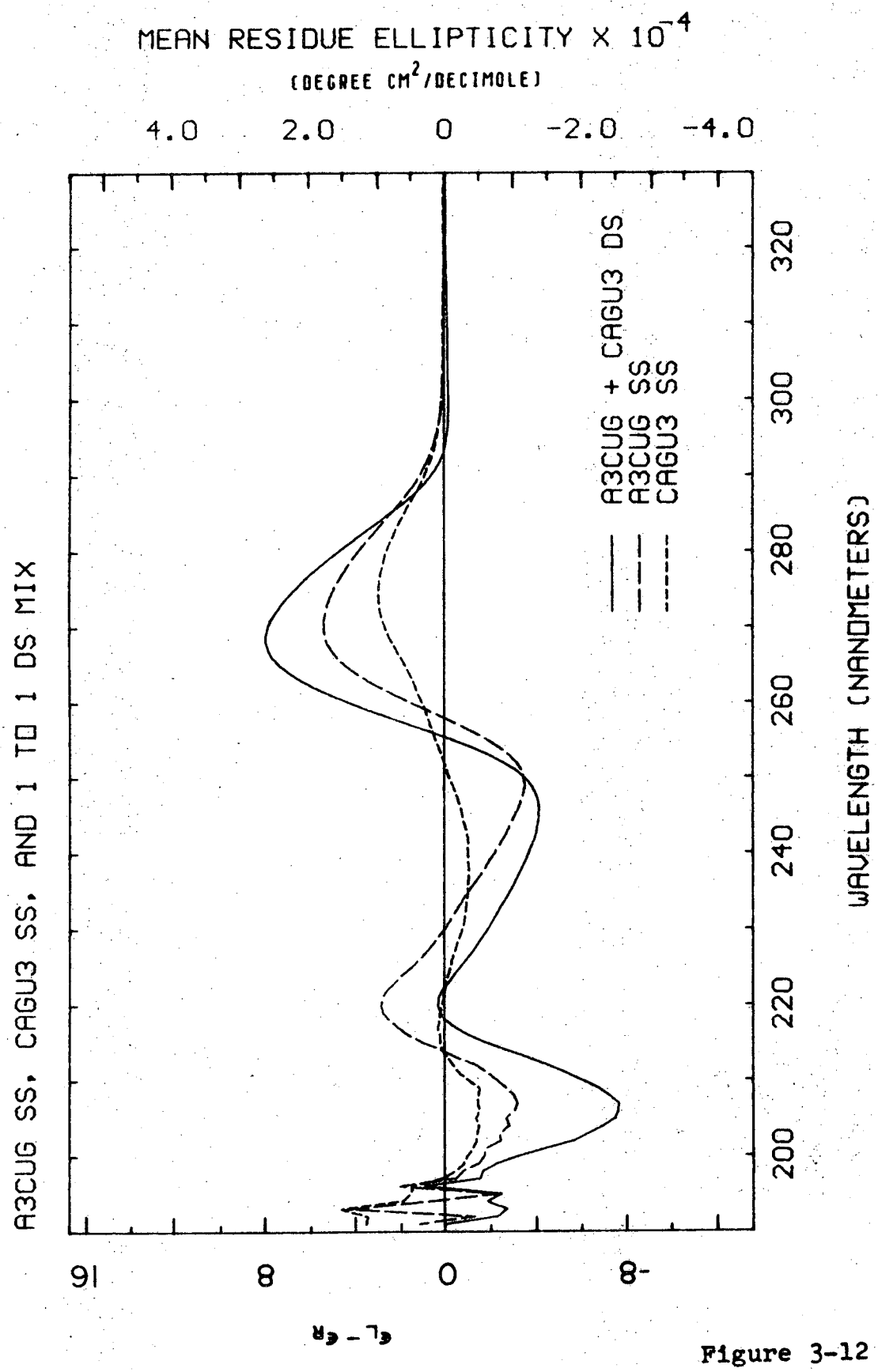


Figure 3-12

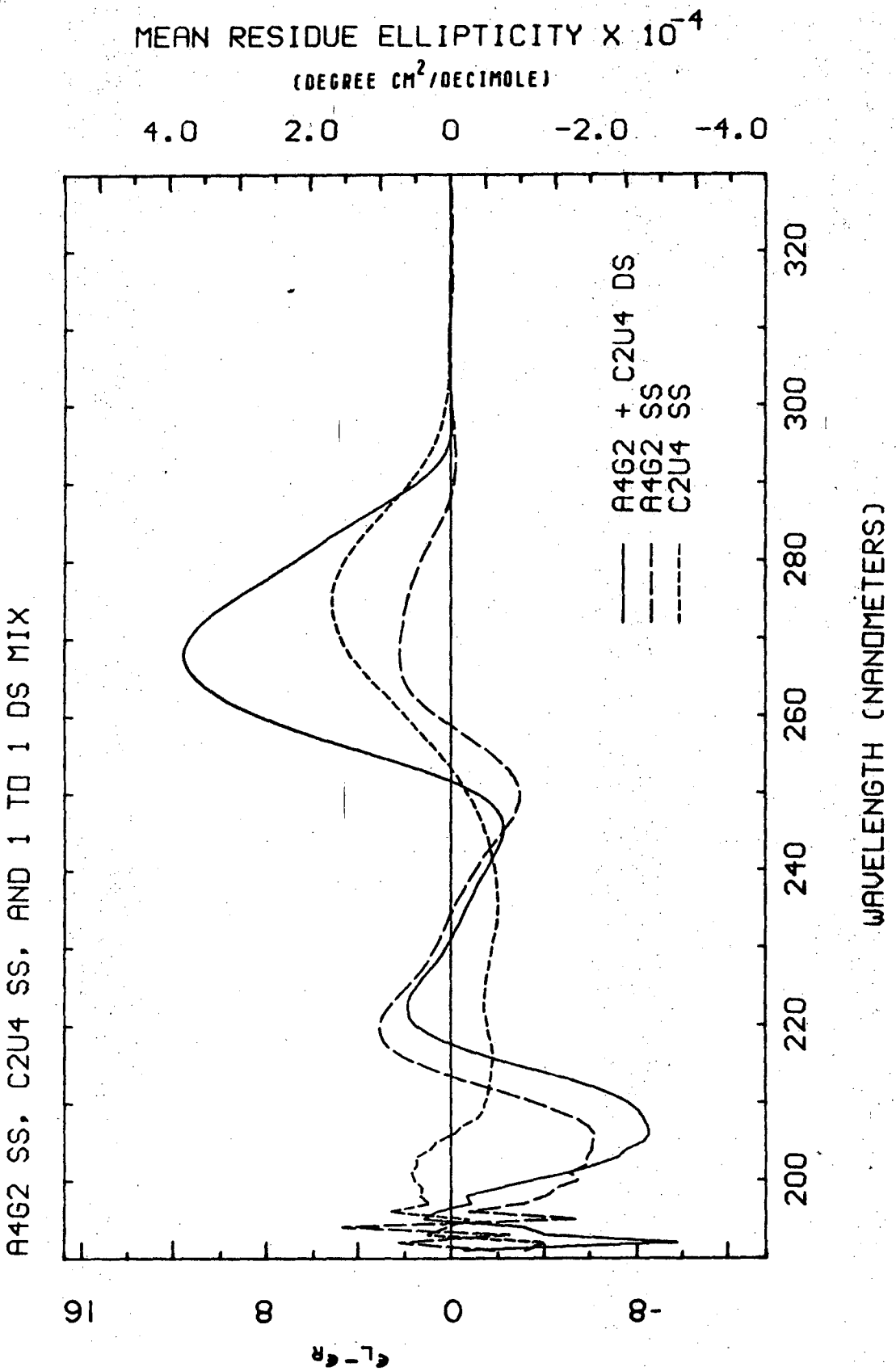

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
4.0
2.0
0
-2.0
-4.0
A4G2 + C2U4 DS
A4G2 SS
C2U4 SS
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
A4G2 SS, C2U4 SS, AND 1 TO 1 DS MIX
16
8
0
-8
$\epsilon_L - \epsilon_R$

Figure 3-13

0 0 0 0 3 9 0 0 7 1 8

pronounced differences due to base composition.

I. $\overrightarrow{A_4CG} + \overrightarrow{CGU_4}$ Have A $(\overrightarrow{CG})_d$ Couple On The End

This molecule is considerably less stable than A_2CGU_2, which has the same kinds of interactions. However, it is still more than 95% double strand at this concentration. Figure 3-14 shows both the single and double strand spectra. There is a small negative band about 300 nm, the 267 nm band is quite low, the 245 nm band is broad but deeper than in A_2CGU_2, and the 210 nm band is broad.

Figure 3-14.

CD of $A_4CG + CGU_4$ double strand ($c_{residue}$ = 0.0090 M, -3.5°C, pH 7, 1 M NaCl) and single strands ($c_{residue} \approx 10^{-4}$ M, 40°C, pH 7, <0.001 M NaCl).

0 0 0 0 3 9 0 0 9 1 9

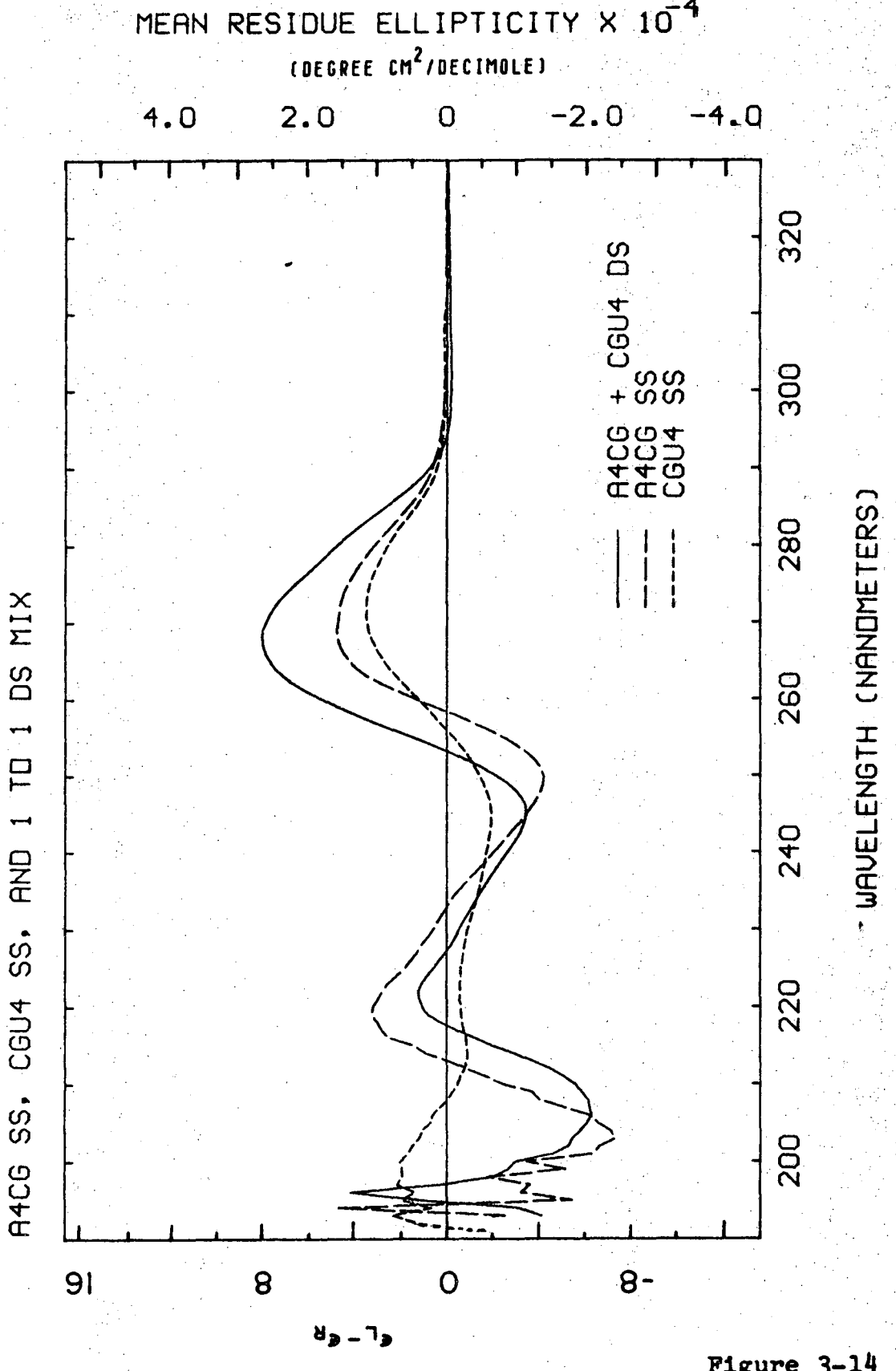


Figure 3-14

CHAPTER IV

Discussion of Results

A. Analysis For Individual First Neighbor Contributions

1. Theory

The first neighbor contributions are present in the oligomers in linear combinations which in some cases makes it difficult to see the effect due to the individual contributions. Given a set of oligomers containing appropriate combinations of all of the contributions, it is in theory a simple process to extract the individual contributions. At a single wavelength the spectral value, S, for an oligomer can be written (see eq. 1-14 and 1-15):

$$(4\text{-}1) \qquad S = \frac{1}{m}\left(\sum_{J=1}^{4} \sum_{K=1}^{L} Q_{JK}[\overrightarrow{JK}] + 1/2 \sum_{i}^{\substack{4 \\ \text{end} \\ \text{bases}}} [\theta]_{\text{monomers}} \right)$$

m = n for single strands with chainlength n, and L = 4. For double strands, m = 2n and we sum over L = 5-J to get the 10 double strand first neighbor terms. We have a measured value for S, $[\theta]_{\text{monomers}}$ and can get the Q_{JK} from the oligomer composition. It is then necessary to solve a sufficient number of such linear equations for the $[\overrightarrow{JK}]$. Of course, a sufficient number is 16 for single strands and 10 for double strands. If the equations are not linearly independent or all the contributions are

0 0 0 0 3 9 0 0 7 2 0

not represented we do not have a complete set and the system of equations cannot be solved.

Let us rewrite eq. (4-1) in matrix form:

$$\underset{\sim}{S} = \underset{\sim}{M}^{-1} \cdot \left(\underset{\sim}{Q} \cdot \underset{\sim}{C} + \underset{\sim}{Y} \cdot \underset{\sim}{E} \right) \tag{4-2}$$

$\underset{\sim}{S}$ is the matrix of measured spectra and is v×t where v is the number of such spectra and t is the number of wavelengths. $\underset{\sim}{Q}$ is the matrix of coefficients of the contributions $\underset{\sim}{C}$ in the oligomers and is simply related to the sequence. $\underset{\sim}{Q}$ is v×ℓ and $\underset{\sim}{C}$ is ℓ × t, where ℓ = 10 for double strands and 16 for single strands. Table 4-1 gives the $\underset{\sim}{Q}$ matrix for the set of double strand oligomers reported in Chapter III. $\underset{\sim}{E}$ is the 4×t array of monomer spectra, row 1 being the spectrum for A, row 2 for C, row 3 for G, and row 4 for U. $\underset{\sim}{Y}$ is the coefficient matrix for the end bases, illustrated in table 4-2 for the double strand oligomers. $\underset{\sim}{M}$ is the v×v matrix with zeroes for off-diagonal elements and values of m for the diagonal elements. m_{11} corresponds to row 1 (spectrum 1) of S, m_{22} to row 2 of S, etc.

We now solve eq. (4-2) for $\underset{\sim}{C}$, the rows of which are the first-neighbor contributions. The elements of $\underset{\sim}{M}^{-1}$ are simply

$$\underset{\sim}{M}^{-1} = \begin{pmatrix} \frac{1}{m_{11}} & & & 0 \\ & \frac{1}{m_{22}} & & \\ & & \ddots & \\ 0 & & & \frac{1}{m_{vv}} \end{pmatrix}, \quad \text{so}$$

Table 4-1. The $\mathcal{Q}$ matrix for double strands.

6.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	FOR A4U4 DS
10.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	FOR A6U6 DS
6.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	1.0	FOR A4UAU4 DS
5.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	FOR A3CU4 + A4GU3 DS
7.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	FOR A5CU4 + A4GU5 DS
2.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	FOR A2GCU2 DS
4.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	FOR A3GCU3 DS
6.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	FOR A4GCU4 DS
8.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	FOR A5GCU5 DS
2.0	2.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	FOR A2CGU2 DS
4.0	2.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	FOR A3CGU3 DS
2.0	0.0	0.0	0.0	0.0	0.0	1.0	2.0	0.0	0.0	FOR U2CGA2 DS
3.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	FOR A4G2 + C2U4 DS
2.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	FOR A3CUG + CAGU3 DS
3.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	FOR A4CG + CGU4 DS

Table 4-2. χ matrix (end coefficients) for double strands.

THESE ARE THE MONOMER COEFFICIENTS FOR THE END BASES A	C	G	U	CHAINLENGTH	
1.0	0.0	0.0	1.0	8.0	FOR A4U4 DS
1.0	0.0	0.0	1.0	12.0	FOR A6U6 DS
1.0	0.0	0.0	1.0	10.0	FOR A4UAU4 DS
1.0	0.0	0.0	1.0	8.0	FOR A3CU4 + A4GU3 DS
1.0	0.0	0.0	1.0	10.0	FOR A5CU4 + A4GU5 DS
1.0	0.0	0.0	1.0	6.0	FOR A2GCU2 DS
1.0	0.0	0.0	1.0	8.0	FOR A3GCU3 DS
1.0	0.0	0.0	1.0	10.0	FOR A4GCU4 DS
1.0	0.0	0.0	1.0	12.0	FOR A5GCU5 DS
1.0	0.0	0.0	1.0	6.0	FOR A2CGU2 DS
1.0	0.0	0.0	1.0	8.0	FOR A3CGU3 DS
1.0	0.0	0.0	1.0	6.0	FOR U2CGA2 DS
.5	.5	.5	.5	6.0	FOR A4G2 + C2U4 DS
.5	.5	.5	.5	6.0	FOR A3CUG + CAGU3 DS
.5	.5	.5	.5	6.0	FOR A4CG + CGU4 DS

$$(4\text{-}3) \qquad \mathfrak{M}\cdot\mathfrak{S} - \mathfrak{X}\cdot\mathfrak{E} = \mathfrak{Q}\cdot\mathfrak{C}$$

$$(4\text{-}4) \qquad \mathfrak{Q}^{t}\cdot\left(\mathfrak{M}\cdot\mathfrak{S} - \mathfrak{X}\cdot\mathfrak{E}\right) = \mathfrak{Q}^{t}\cdot\mathfrak{Q}\cdot\mathfrak{C}$$

$\mathfrak{Q}^{t}$ is the transpose of $\mathfrak{Q}$ and $\mathfrak{Q}^{t}\cdot\mathfrak{Q}$ is $\ell \times \ell$ square symmetric and has an inverse provided the set of oligomers in S spans the ℓ-dimensional sequence space. On obtaining the inverse by standard computational methods we have the solution:

$$(4\text{-}5) \qquad \left(\mathfrak{Q}^{t}\cdot\mathfrak{Q}\right)^{-1}\cdot\mathfrak{Q}^{t}\cdot\left(\mathfrak{M}\cdot\mathfrak{S} - \mathfrak{X}\cdot\mathfrak{E}\right) = \mathfrak{C}\ .$$

The Fortran coded program, CDCONT, described in the appendix accomplished this procedure, recalculates the oligomer spectra by eq. (4-2), and calculates CD spectra of other RNA's with known sequence or, ignoring end effects, with known nearest neighbor frequencies.

2. Double Strand CD Contributions

The contributions extracted according to the above description are collected in figures 4-1 through 4-4. Since they are in reality difference spectra their irregular character is not too surprising. They were not smoothed as it was felt that this would introduce an additional error into subsequent calculations. There is little that can be said generally about them since they all have unique shapes. The contributions leading to a negative band at ~300 nm are: $[\overrightarrow{CC}]_d$, $[\overrightarrow{GC}]_d$, $[\overrightarrow{AC}]_d$, and $[\overrightarrow{GA}]_d$, and to a small extent $[\overrightarrow{AU}]_d$. On the contrary, $[\overrightarrow{CG}]_d$, $[\overrightarrow{AG}]_d$, and possibly $[\overrightarrow{CA}]_d$ and $[\overrightarrow{UA}]_d$

0 0 0 0 5 9 0 0 9 2 2

Figure 4-1.

The double strand first neighbor CD contributions $[\overrightarrow{AA}]_d$, $[\overrightarrow{AU}]_d$, and $[\overrightarrow{UA}]_d$.

Figure 4-2.

The double strand first neighbor CD contributions $[\overrightarrow{CC}]_d$, $[\overrightarrow{CG}]_d$, and $[\overrightarrow{GC}]_d$.

Figure 4-3.

The double strand first neighbor CD contributions $[\overrightarrow{AC}]_d$ and $[\overrightarrow{CA}]_d$.

Figure 4-4.

The double strand first neighbor CD contributions $[\overrightarrow{AG}]_d$ and $[\overrightarrow{GA}]_d$.

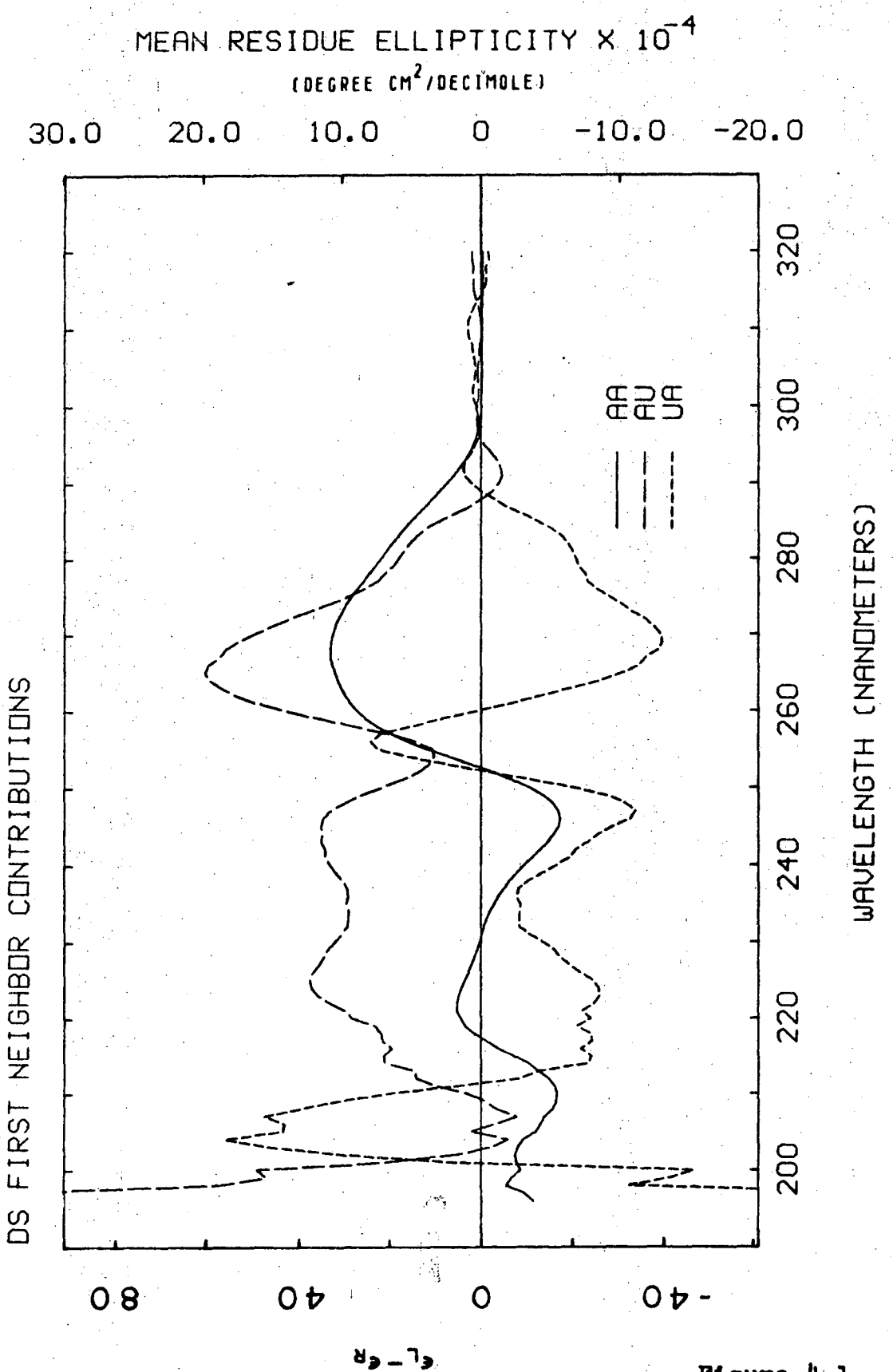

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
30.0
20.0
10.0
0
-10.0
-20.0
AA
AU
UA
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
DS FIRST NEIGHBOR CONTRIBUTIONS
80
40
0
-40
$\epsilon_L - \epsilon_R$

Figure 4-1

0 0 0 0 3 9 0 0 0 2 3

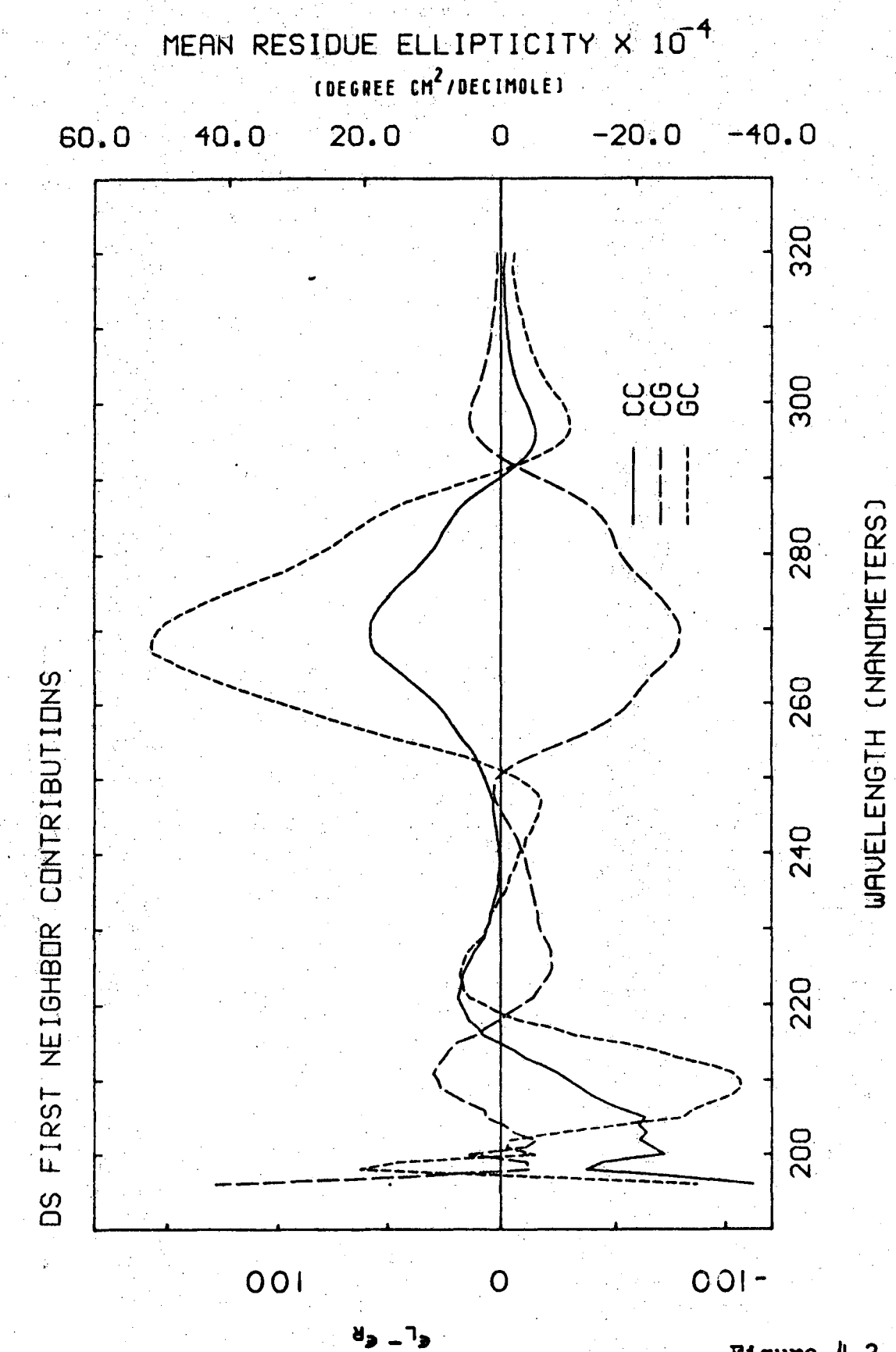


Figure 4-2

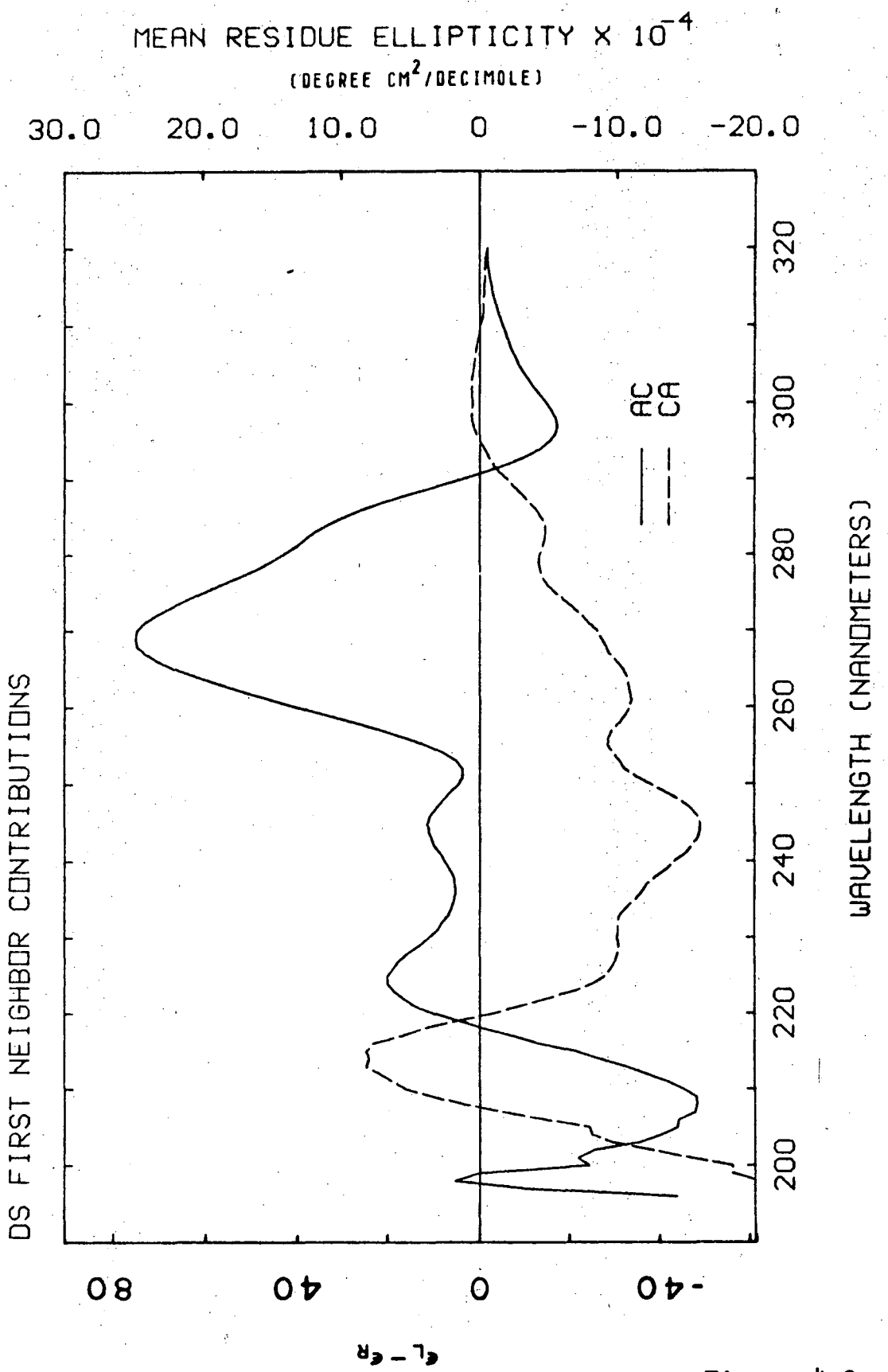

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
30.0
20.0
10.0
0
-10.0
-20.0
AC
CA
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
DS FIRST NEIGHBOR CONTRIBUTIONS
80
40
0
-40
$\epsilon_L - \epsilon_R$

Figure 4-3

0 0 0 0 3 9 0 0 9 2 4

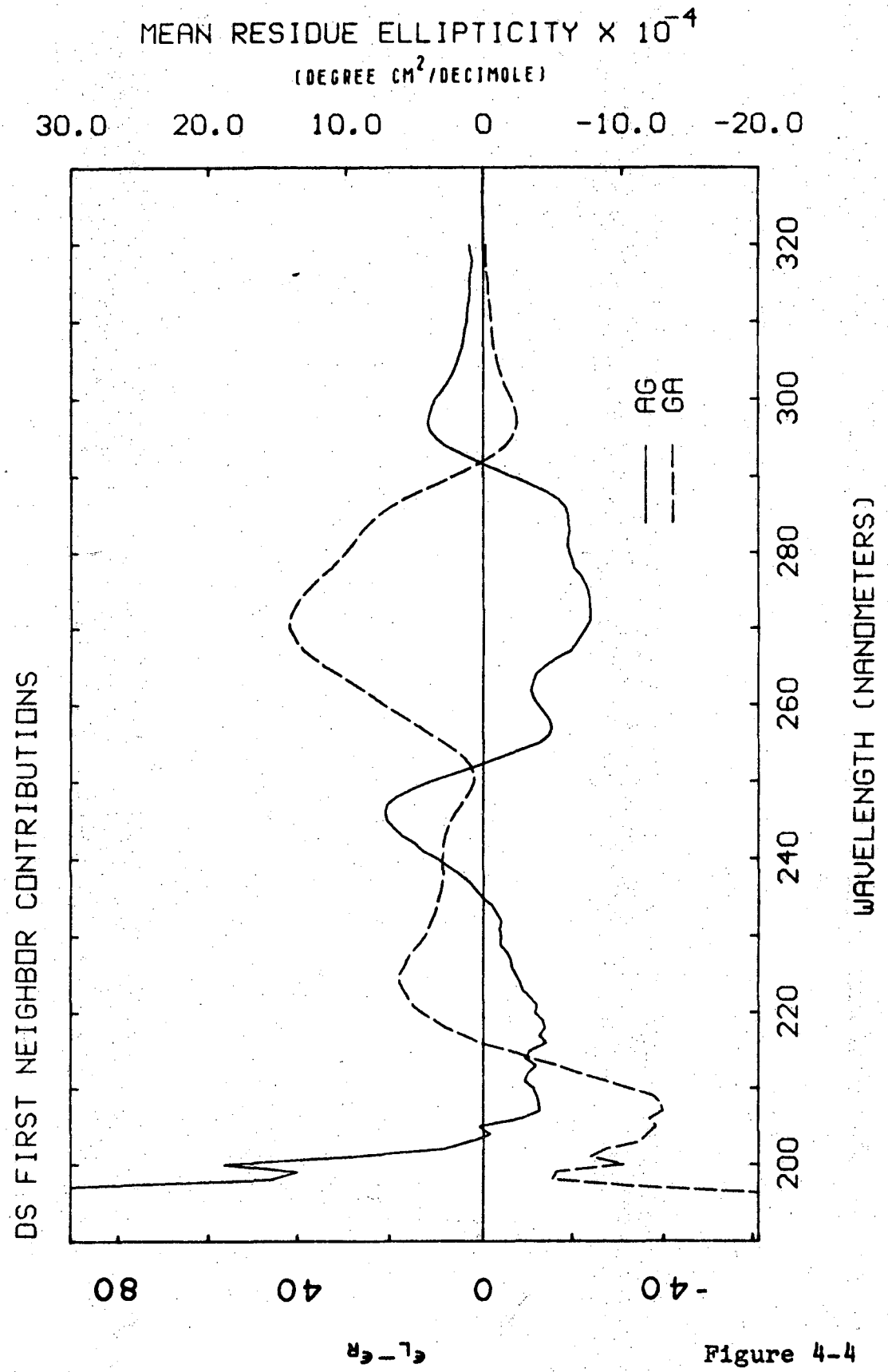


Figure 4-4

contribute a positive band in the long wavelength region. They all provide some positive or negative contribution at 267 nm accounting for the variations in magnitude there. The $[\overrightarrow{GC}]_d$ contributes a very large magnitude here. Also, they all show some kind of shoulder on the red side of this peak, indicating either a common electronic feature or the inability of this procedure to sort out the couples to which this feature belongs. The $[\overrightarrow{AA}]_d$, $[\overrightarrow{UA}]_d$, $[\overrightarrow{GC}]_d$, $[\overrightarrow{CA}]_d$, and $[\overrightarrow{GA}]_d$ contribute magnitude (negative CD) to the 245 nm trough while $[\overrightarrow{AU}]_d$, $[\overrightarrow{AC}]_d$, and $[\overrightarrow{AG}]_d$ contribute positively. The 220 nm peak gets magnitude from $[\overrightarrow{AA}]$, $[\overrightarrow{AU}]$, $[\overrightarrow{CC}]$, $[\overrightarrow{GC}]$, $[\overrightarrow{AC}]$, and $[\overrightarrow{GA}]$, while all the others reduce it. Most of the couples give the 210 nm trough, $[\overrightarrow{GC}]$ and $[\overrightarrow{AC}]$ being the largest. $[\overrightarrow{UA}]$, $[\overrightarrow{CG}]$, and $[\overrightarrow{CA}]$ give positive contributions here.

a. Linear Combinations of Spectra Giving the Double Strand Contributions

In an effort to get a better understanding of the significance of these results we present in table 4-3 the linear combinations of spectra made by the program to calculate each contribution. (Actually, linear combinations of the rows of the matrix $\left(\underline{\underline{M}}\cdot\underline{\underline{S}} - \underline{\underline{X}}\cdot\underline{\underline{E}}\right)$, the monomer and chainlength corrected spectra.) In looking at this table we see that fortunately the $A_3CUG + CAGU_3$, and U_2CGA_2 spectra are involved only in the calculation of the contributions $[\overrightarrow{CA}]_d$ and $[\overrightarrow{GA}]_d$ respectively so any

Table 4-3. Coefficients for linear combinations of measured double strand spectra composing the double strand contributions.

	A_4U_4	A_6U_6	A_4UAU_4	A_3CU_4 + A_4GU_3	A_5CU_4 + A_4GU_5	A_2GCU_2	A_3GCU_3	A_4GCU_4	A_5GCU_5	A_2CGU_2	A_3CGU_3	U_2CGA_2	A_4G_2 + C_2U_4	A_3CUG + $CAGU_3$	A_4CG + CGU_4	Sum of absolute values
[$\overline{AA}$]	-.0625	.0625	0.0000	-.0313	.0312	-.0937	-.0312	.0313	.0938	-.0313	.0312	.0000	.0000	.0000	-.0000	0.5
[$\overline{AC}$]	.0000	-.0000	-.0000	-.0000	-.0000	0.0000	0.0000	0.0000	0.0000	.5000	.5000	0.0000	0.0000	-.0000	-1.0000	2.0
[$\overline{AG}$]	.3750	-.3750	-.0000	.6875	.3125	.5625	.1875	-.1875	-.5625	-.3125	-.6875	0.0000	0.0000	.0000	1.0000	5.25
[$\overline{AU}$]	1.0000	0.0000	-.0000	.2500	-.2500	.7500	.2500	-.2500	-.7500	.2500	-.2500	-.0000	-.0000	-.0000	.0000	4.0
[$\overline{CA}$]	-.2500	.2500	0.0000	-.6250	-.3750	-.3750	-.1250	.1250	.3750	-.1250	.1250	.0000	.0000	1.0000	-.0000	3.75
[$\overline{CC}$]	-.1875	.1875	.0000	-.5938	-.4063	-.2813	-.0938	.0938	.2813	.4062	.5937	0.0000	1.0000	-.0000	-1.0000	4.125
[$\overline{CG}$]	.1875	-.1875	-.0000	.0938	-.0937	.2813	.0938	-.0938	-.2813	-.4062	-.5937	0.0000	0.0000	.0000	2.0000	4.313
[$\overline{GA}$]	-.0313	.0312	0.0000	-.0156	.0156	-.0469	-.0156	.0156	.0469	.2344	.2656	.5000	0.0000	-.0000	-1.0000	2.219
[$\overline{GC}$]	-.4375	.4375	0.0000	-1.2188	-.7813	-.4063	.0313	.4688	.9063	.7812	1.2187	0.0000	0.0000	-.0000	-2.0000	8.688
[$\overline{UA}$]	-1.6250	-.3750	1.0000	-.3125	.3125	-.9375	-.3125	.3125	.9375	-.3125	.3125	-.0000	-.0000	-.0000	-.0000	5.75

spectral errors are not transmitted through to any of the other calculated contributions. Of course, the same is also true with $A_4G_2 + C_2U_4$. On the other side of the coin, the A_4U_4 and $A_4CG + CGU_4$ spectra enter with unusually large coefficients in many contributions. This is bad because both probably contain some single strands at the concentrations used. For the sake of estimating the error, let us say that each measured spectrum has the same error. Then that error will be included in the error in the contribution according to the absolute value of its coefficient in the linear combination. Looking at the final column of table 4-3 we see the sum of the coefficients making up the interactions. The $[\overrightarrow{AA}]_d$ contribution should have error bars equal to half those characteristic for the measured spectra; we have essentially performed an averaging procedure for this contribution. The $[\overrightarrow{AC}]_d$ contribution however has twice the normal error in a spectrum; an $[\overrightarrow{AU}]_d$, 4 times; and a $[\overrightarrow{GC}]_d$, 8.7 times the normal uncertainty. If the experimental uncertainty is ±3%, then the $[\overrightarrow{GC}]_d$ is correct to about ±25%. Of course, that is the worst case but it does put many of the little wiggles in the previous figures in a different light.

b. Reconstituting the Input Spectra from the $[\overrightarrow{JK}]_d$

The input spectra can be recalculated from the interactions and the $\mathfrak{A}$ matrix according to eq. (4-2). The

0 0 0 0 3 9 0 0 7 2 6

general result is that there is good but not exact agreement between the recalculated and measured spectra for cases such as the A_nU_n series illustrated in Figure 4-5. Here, there are more than enough oligomers to extract the necessary contributions and spectral differences get averaged out in the recalculation. That is to say, the calculated $[\overrightarrow{AA}]_d$ and $[\overrightarrow{AU}]_d$ contributions are the ones that best fit both A_4U_4, and A_6U_6, and the other oligomers, simultaneously. The differences are due to experimental error and the failure of the first-neighbor approximation to completely explain oligomer properties.

In cases such as A_4UAU_4, the recalculated spectrum fits the measured exactly. This is because there is only one oligomer involving the $[\overrightarrow{UA}]_d$ interaction. Therefore, the linear combinations of oligomer spectra throw all of the difference between the appropriate combination of $[\overrightarrow{AA}]_d$, $[\overrightarrow{AU}]_d$ into the $[\overrightarrow{UA}]_d$ contribution. The other calculated spectra which fit the measured spectra exactly are U_2CGA_2, $A_3CUG + CAGU_3$, $A_4G_2 + C_2U_4$, and $A_4CG + CGU_4$. If one of these spectra is left out of the $\underset{\sim}{S}$ array, the matrix of $(\underset{\sim}{Q}^t \cdot \underset{\sim}{Q})^{-1}$ becomes singular and the equation cannot be solved for $\underset{\sim}{C}$.

c. Predicting Other Double Strand Spectra

The usefulness of this analysis depends on its predictive capacities. Can we predict the spectra of other RNA's of known sequence? As a test, we calculate

Figure 4-5.

Measured spectra for A_4U_4 and A_6U_6 and spectra recalculated from appropriate combinations of $[\overrightarrow{AA}]_d$ and $[\overrightarrow{AU}]_d$.

0 0 0 0 3 9 0 0 7 2 7

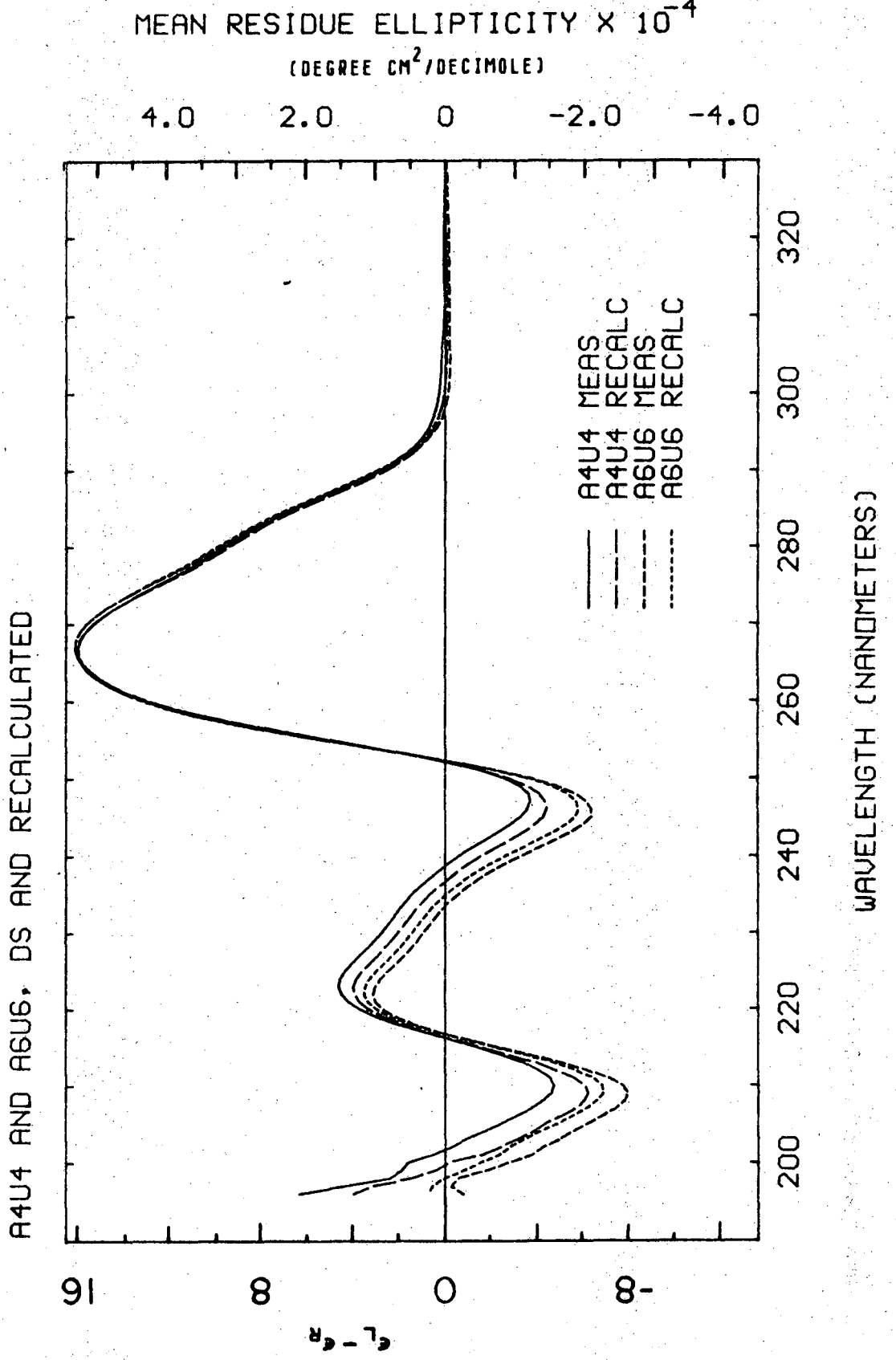


Figure 4-5

the CD spectra of several simple sequence synthetic RNA polymers. The results are presented in figures 4-6 through 4-10. With the exception of poly C:G, all give good qualitative agreement in their shapes. For instance, the calculated poly CG:CG curve in figure 4-9 predicts a long wavelength minimum, followed by a maximum, then a minimum, and further low wavelength extrema that are present in both the measured and calculated spectra. The agreement for the crossovers and the magnitude of the long wavelength band leaves something to be desired, however. This can be due to several causes. First, the calculated values are obtained from differences between large numbers, leading to a large uncertainty; the exact shapes and magnitudes of the contributions cannot be determined. This means that the existence of an extremum in a spectral region is more accurately predicted than the exact shape or magnitude. Therefore, we consider "good" agreement more as a correlation in shapes rather than absolute position and magnitude. This problem is no doubt a major contributor to the discrepancy between the measured and calculated poly AC:GU spectra. The calculation is made by summing equal parts of the $[\overrightarrow{AC}]_d$ and $[\overrightarrow{CA}]_d$ contributions. The $[\overrightarrow{CA}]_d$ contribution is not well characterized at this time. The second source of disagreement is simply that the measured polymer spectra might be in error. This is a fair possibility in the poly C:G case since there are usually problems with poly

0 0 0 0 3 9 0 0 7 2 8

Figure 4-6.

CD of poly A:U, measured (————), 1/2 $[\overrightarrow{\overline{AA}}]_d$ (--------).

Figure 4-7.

CD of poly AU:AU, measured (————), 1/4 $[\overrightarrow{\overline{AU}}]_d$ + 1/4 $[\overrightarrow{\overline{UA}}]_d$ (--------).

Figure 4-8.

CD of poly C:G, measured (————), 1/2 $[\overrightarrow{\overline{CC}}]_d$ (--------).

Figure 4-9.

CD of poly CG:CG, measured (————), 1/4 $[\overrightarrow{\overline{CG}}]_d$ + 1/4 $[\overrightarrow{\overline{GC}}]_d$ (--------).

Figure 4-10.

CD of poly AC:UG, measured (————), 1/4 $[\overrightarrow{\overline{AC}}]_d$ + 1/4 $[\overrightarrow{\overline{CA}}]_d$ (--------).

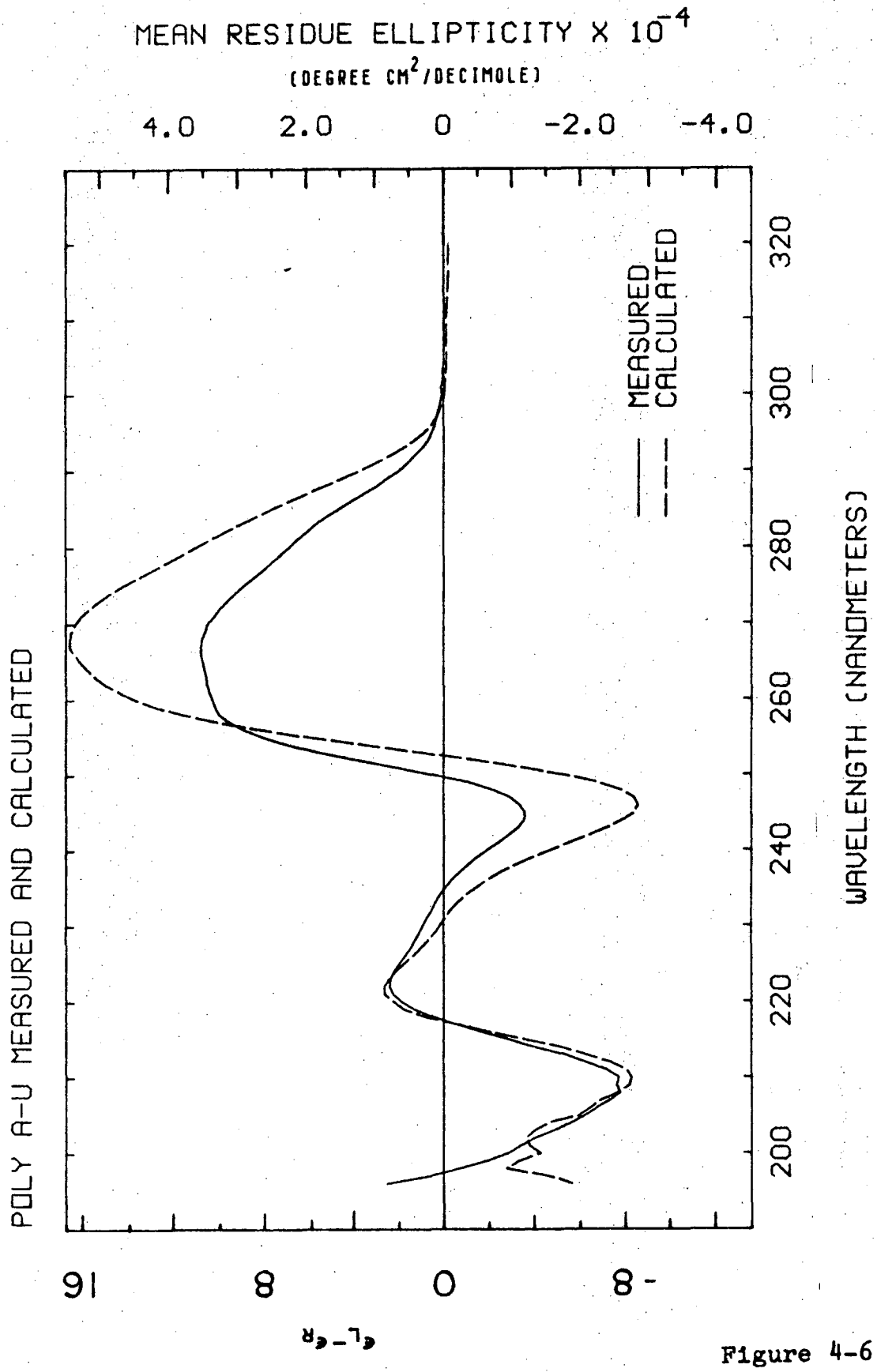

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
4.0
2.0
0
-2.0
-4.0
MEASURED
CALCULATED
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
POLY A-U MEASURED AND CALCULATED
16
8
0
-8
$\epsilon_L - \epsilon_R$

Figure 4-6

0 0 0 0 3 9 0 0 7 2 9

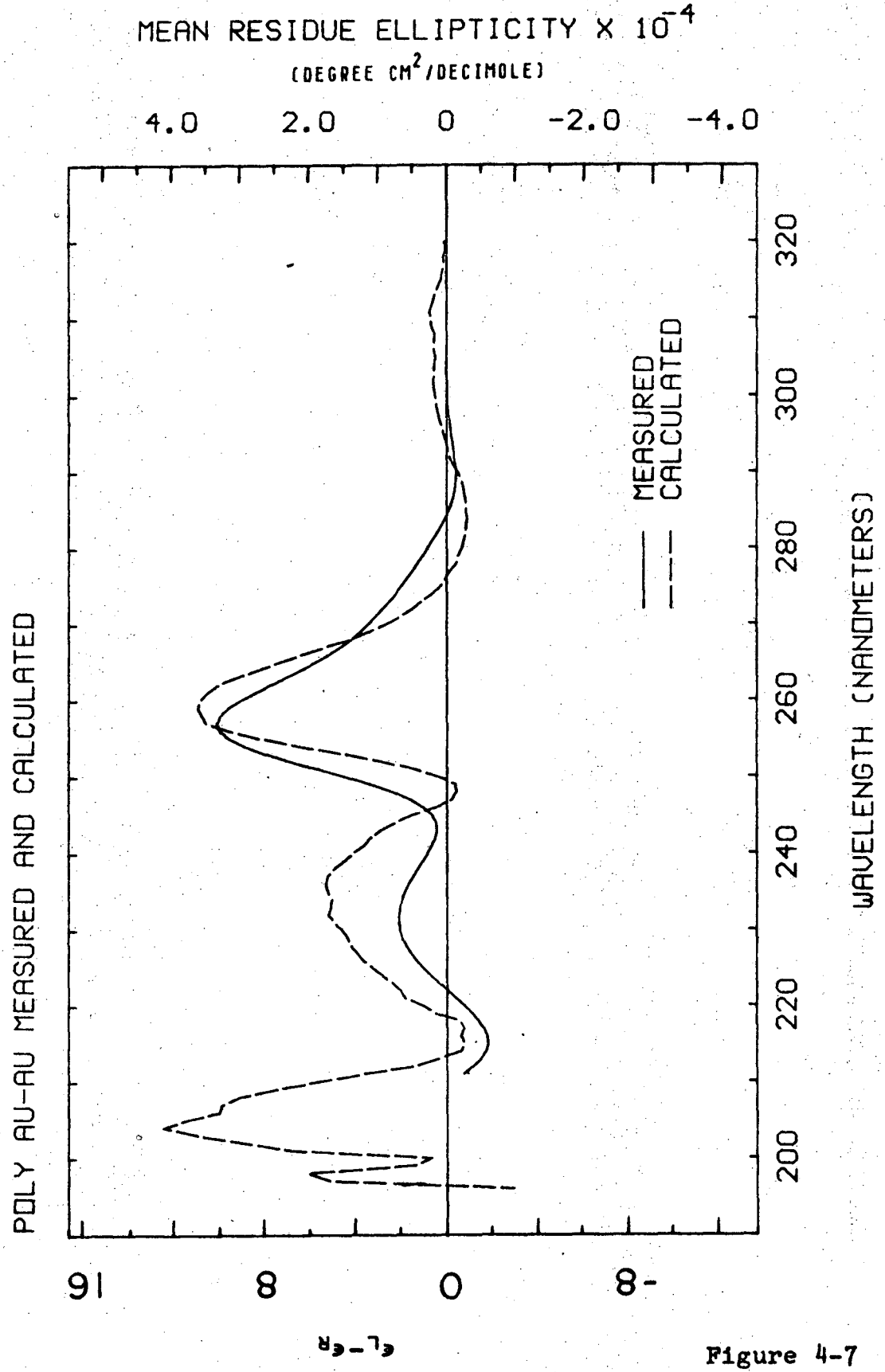


Figure 4-7

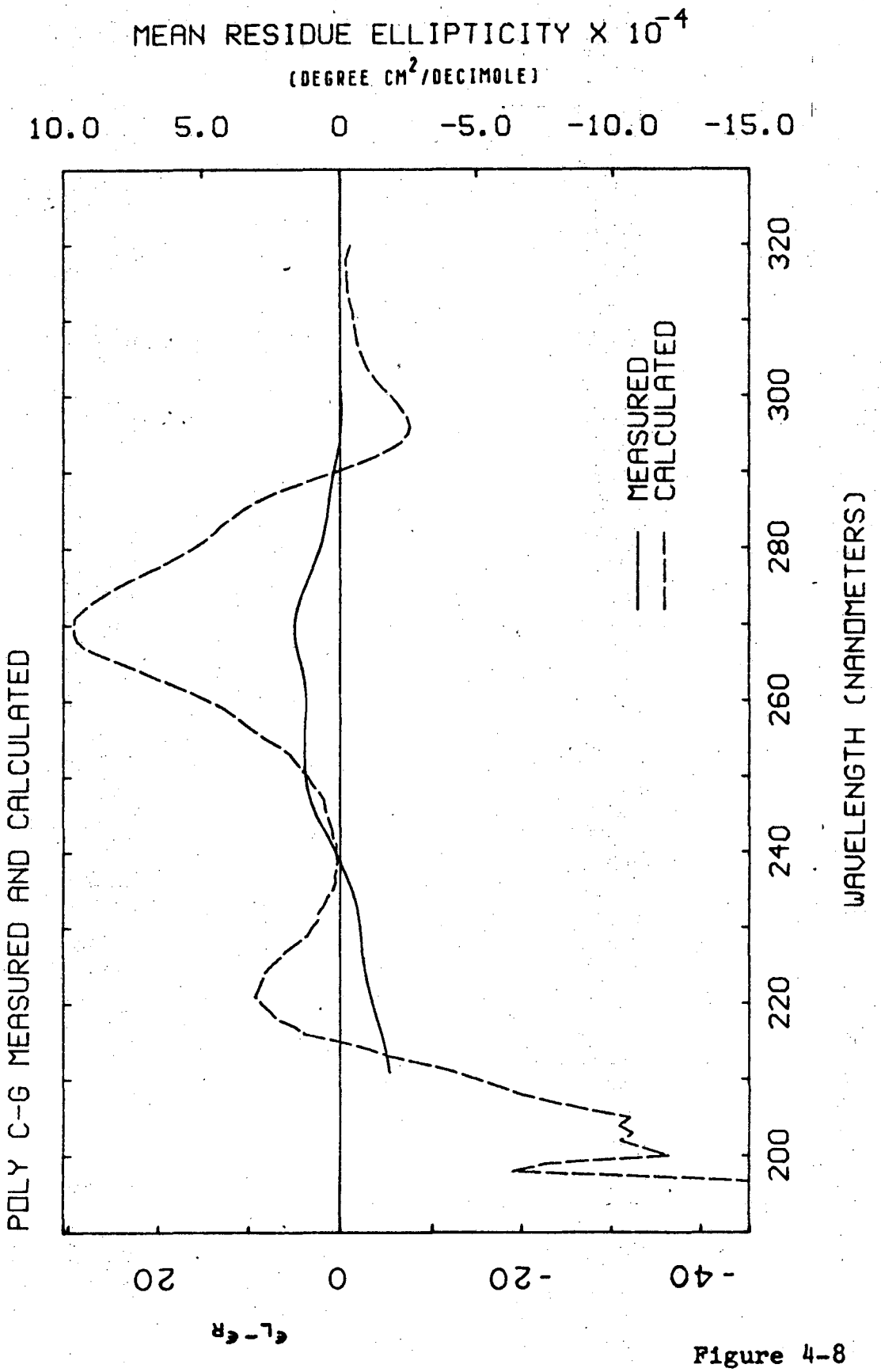

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
10.0
5.0
0
-5.0
-10.0
-15.0
MEASURED
CALCULATED
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
POLY C-G MEASURED AND CALCULATED
20
0
-20
-40
$\epsilon_L - \epsilon_R$

Figure 4-8

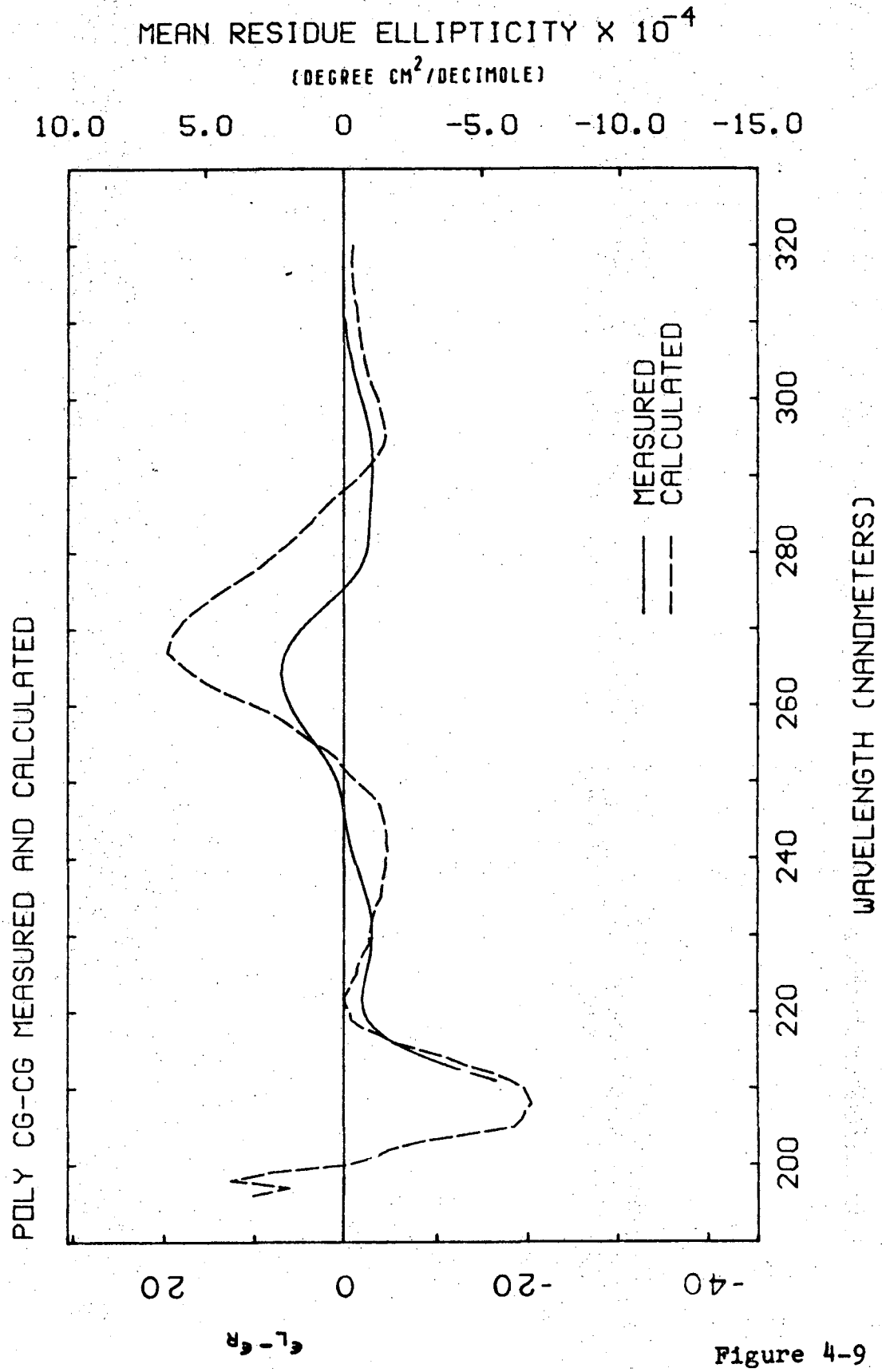

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
10.0
5.0
0
-5.0
-10.0
-15.0
MEASURED
CALCULATED
POLY CG-CG MEASURED AND CALCULATED
$\epsilon_L - \epsilon_R$
20
0
-20
-40
200
220
240
260
280
300
320
WAVELENGTH (NANOMETERS)

Figure 4-9

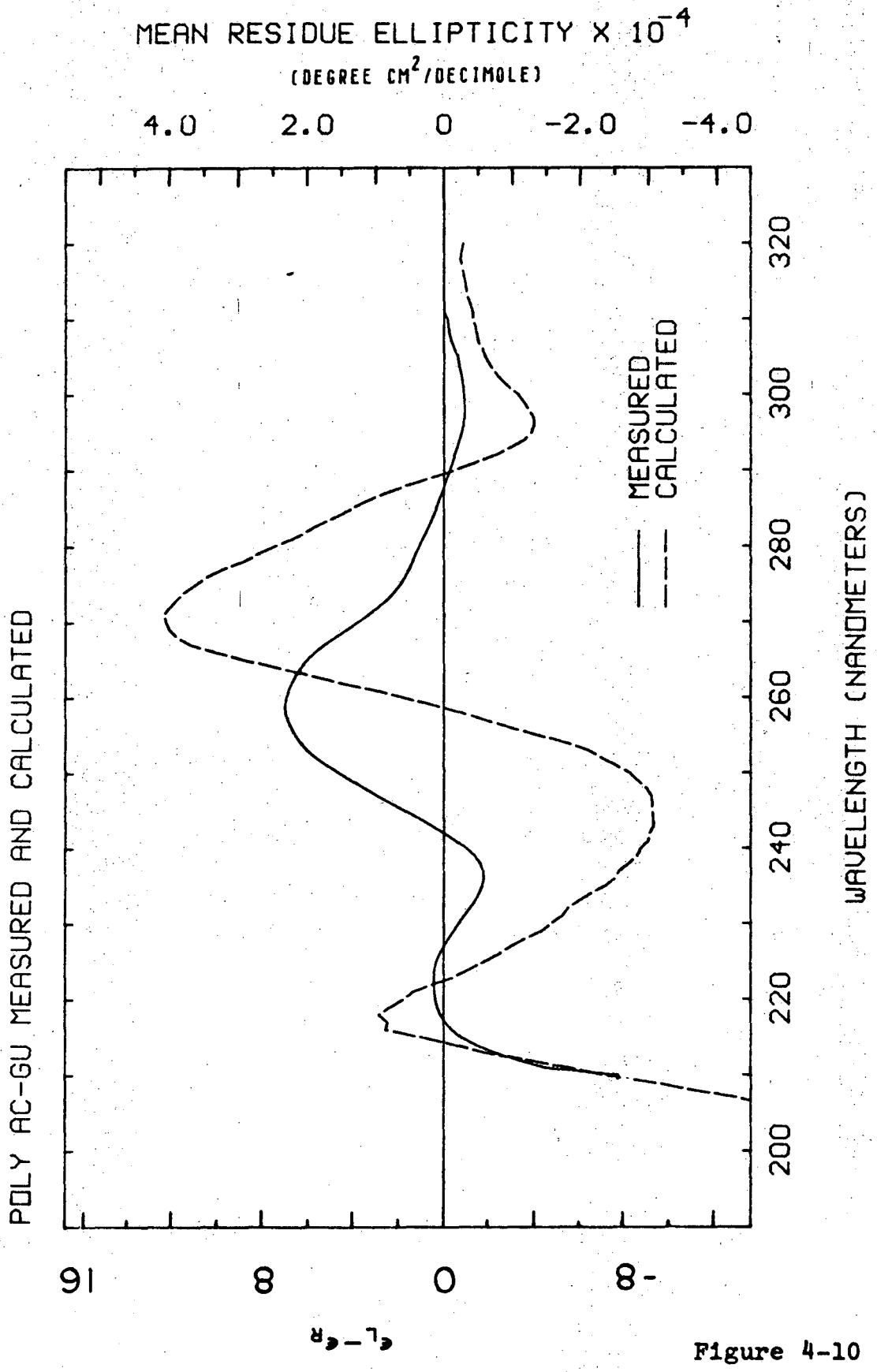

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
4.0
2.0
0
-2.0
-4.0
MEASURED
CALCULATED
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
POLY AC-GU MEASURED AND CALCULATED
16
8
0
-8
$\epsilon_L - \epsilon_R$

Figure 4-10

0 0 0 0 3 9 0 0 7 3 1

G self-aggregation in these systems. A third error is that the double strand couples may exist in a considerably different geometry in a simple sequence polymer than when it is in a chain with varied numbers and kinds of couples. There is ample precedent for this in DNA's which have all purines on one strand and all pyrimidines on the other.[1] However, RNA has not been found in different structural forms as has DNA. Finally, differences may arise from longer range interactions than are accounted for by first neighbor terms.

Theoreticians have predicted[2] that the first neighbor approximation would work poorly for double strand RNA's where the bases have a large degree of tilt with respect to the helix axis, and are located farther from the helix axis than in DNA. The prediction was that this arrangement would lead to a situation where the first neighbor terms would account for less than half of the total CD. Since the measured polymers have such distinctive shapes it is difficult to accept this proposition in the light of the agreement between the calculated and measured CD presented here. If longer range interactions were very important they might considerably alter the shapes of the calculated spectra, leading to a less satisfactory agreement with measurements.

Figure 4-11 presents a calculated poly AG:CU spectrum. This polymer has not been synthesized so we have no check as yet.

Figure 4-11.

Calculated CD for poly AG:CU = 1/4 $[\overrightarrow{AG}]_d$ + 1/4 $[\overrightarrow{GA}]_d$.

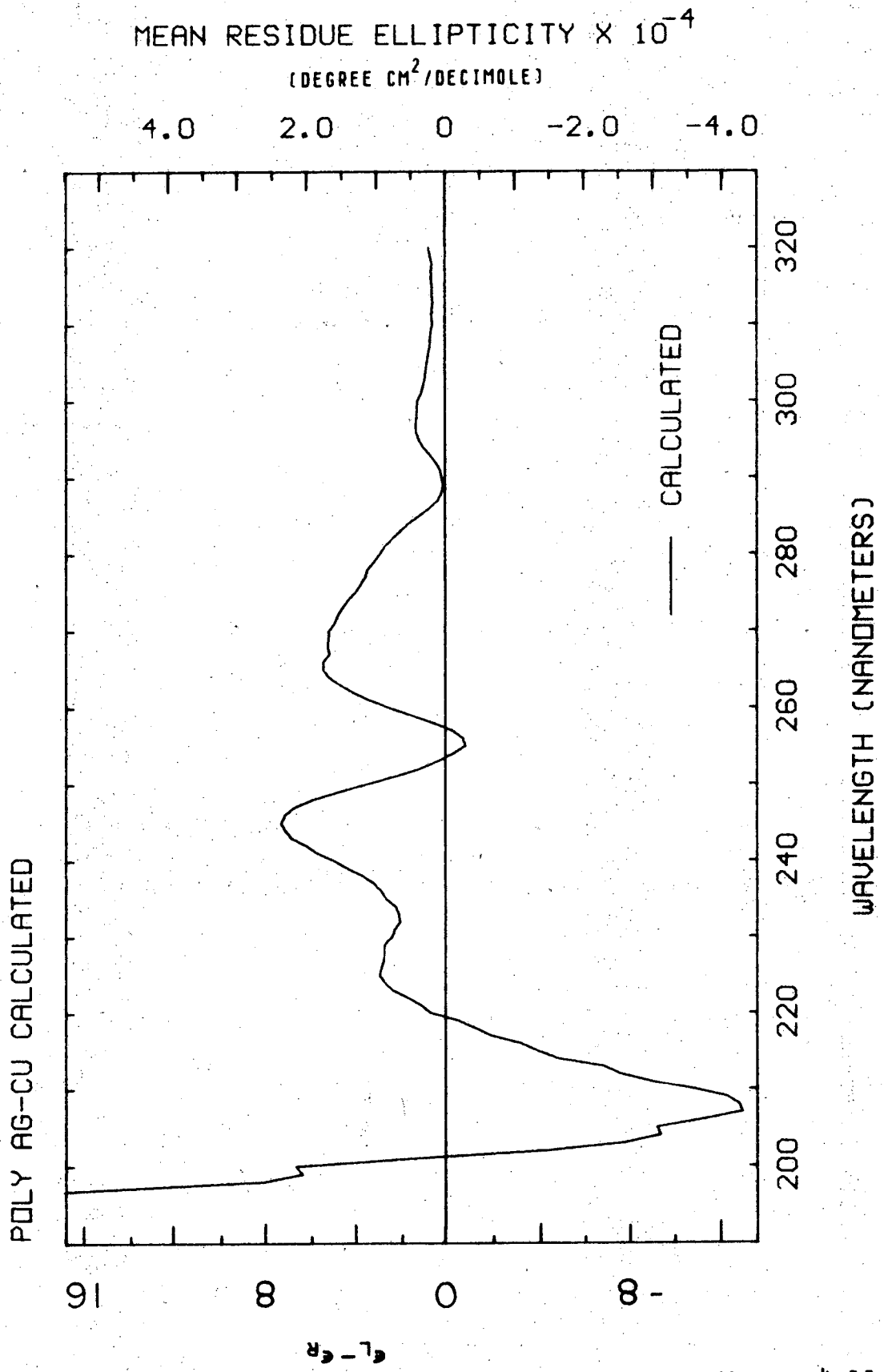

MEAN RESIDUE ELLIPTICITY X 10^{-4}
(DEGREE CM^2/DECIMOLE)
4.0
2.0
0
-2.0
-4.0
CALCULATED
320
300
280
260
240
220
200
WAVELENGTH (NANOMETERS)
POLY AG-CU CALCULATED
16
8
0
-8
$\epsilon_L - \epsilon_R$

Figure 4-11

These polymers probably do not represent the best check possible of the applicability of these contributions toward calculating the CD of other RNA systems. Their unique repeating sequences make it more likely that relatively weak long-range interactions might add in phase to produce aberrations that would not be present in more nearly random sequences. We do not have at present the CD of a double strand RNA (e.g. the replicative form of a viral RNA) for which nearest neighbor frequencies are known. tRNA is also not a very good test case since it has considerable single strand sections, hairpin loops, and a large proportion of unusual bases. The contributions obtained from the double strand oligomers will probably do a better job of fitting the tRNA's than the polymer spectra by Blum's method.[3] She notes that the polymers have a magnitude too low to adequately describe the double strand component of a tRNA CD spectrum. Since the contributions obtained in this work calculate a higher magnitude for the polymers, they will also probably better fit the tRNA, and be better models in general for short double helical RNA.

3. Single Strand CD Contributions

The set of 21 single strand oligomer CD spectra shown in Chapter III do not span the 16-dimensional first neighbor sequence space. That is, the determinant of the $\mathfrak{Q}^t \cdot \mathfrak{Q}$ matrix is zero and the inverse cannot be com-

0 0 0 0 3 9 0 0 7 3 3

puted. Adding other properly chosen spectra would remove the linear dependence and allow single strand contributions to be calculated. Alternatively, one could probably find a subset of sequences (leaving out some of the contributions) for which the 21 spectra are a complete set. It was felt, however, that this is a subject which could be easily deferred to a later time since the dinucleoside phosphate data[4] are quite adequate for calculating most single strand spectra.

B. Suggestions for Further Work

Additional spectra should be measured on other sequences to more closely define the double strand contributions. Particularly molecules involving $(\overrightarrow{CA})_d$ couples need to be included in the set. It would also be helpful to better determine the $[\overrightarrow{UA}]_d$ contribution and the others represented only once in the library of spectra. The single strand story should also be completed in order to take full advantage of Blum's tRNA data.[3]

There are also other methods by which the contributions might be more accurately extracted. A linear programming analysis similar to that of Allen[1] might be employed. Also, recently workers in electron microscopy have devised theoretical constructs for taking into account experimental errors and overlapping pieces of information in order to extract a maximum of structural information.[5] This analysis might be extended to data

such as are presented in this thesis.

0 0 0 0 3 9 0 0 2 3 4

REFERENCES FOR CHAPTER IV.

1. Allen, F.S., Gray, D.M., Roberts, G.P., and Tinoco, Jr., I., Biopolymers, 11, 853 (1972).
2. Johnson, Jr., W.C. and Tinoco, Jr., I., Biopolymers, 7, 727 (1969).
3. Blum, A.D., Ph.D. thesis, U. of Calif. Berkeley (1971).
4. Warshaw, M.M. and Cantor, C.R., Biopolymers, 9, 1079 (1970).
5. Klug, A. and Crowther, R.A., Nature, 238, 435 (1972).

THE STABILITY OF RNA HAIRPIN LOOPS: $A_6C_mU_6$ †

Olke C. Uhlenbeck*, Philip N. Borer, Barbara Dengler, and Ignacio Tinoco, Jr.
Department of Chemistry
University of California
Berkeley, California 94720

*Present Address:

Department of Biochemistry
University of Illinois
Urbana, Illinois 61801

† Journal of Molecular Biology (in press).

0 0 3 0 3 9 0 0 7 3 5

CHAPTER V

The Stability of RNA Hairpin Loops: $A_6C_mU_6$

Abstract

The thermodynamics and circular dichroism of a series of $A_6C_mU_6$ (m = 4,5,6,8) oligoribonucleotides have been studied. These molecules from intramolecular hairpin loops at low temperatures and therefore are useful models for similar structures which occur in larger, natural RNA molecules. The stability of the helix forming the stem of these loops was found to be considerably greater than an intermolecular helix with the same length and composition. The most stable loop is m = 6. The enthalpy for initiation of the loop is unfavorable; it ranges from +24 kcal for m = 4 to +21 kcal for m = 6. The maximum in stability for the C_6 loop and the large positive enthalpy for loop initiation are in disagreement with expectations from simple theories assuming a Gaussian distribution of end to end distances. Loop strain for m = 4 and 5 and the unstacking of the cytosines on loop formation are likely physical explanations for these thermodynamic data. The circular dichroism spectrum of cytosine residues in the C_6 and C_8 loops is very similar to the spectrum of single stranded oligoribocytidylate. However, the cytosine

0 0 0 0 3 9 0 0 7 3 6

residues in the C_5 loop have a very different CD spectrum from the corresponding oligo C_5 spectrum. In accord with the thermodynamic data, we conclude from the circular dichroism data that the C_5 loop has an altered conformation from the C_6 and C_8 loops.

A. Introduction

Hairpin loops are a necessary feature of RNA secondary structure. Whenever an antiparallel helical region is formed by base pairing between complementary sequences of an RNA strand, a hairpin loop must be formed as well. One expects the loop to stabilize its intramolecular helical region relative to an identical intermolecular helix. However, in order to accurately predict a secondary structure on the basis of sequence (Kallenbach, 1968; Tinoco, Uhlenbeck, and Levine, 1971; DeLisi and Crothers, 1971) it is necessary to have precise thermodynamic information about loop stability. Very little thermodynamic data on loops is presently available. Scheffler et al. (1970) have studied loops formed by deoxy AT oligomers of various lengths. Loops have also been obtained as fragments of tRNA molecules (Romer, et al., 1969), Coutts, 1971)). It has been difficult to determine from these data (1) the amount which a loop stabilizes its helix and (2) the effect of the number of residues in the loop on the stability

of the helix.

In this paper, RNA oligomers of the form $A_6C_mU_6$ (m = 4,5,6 and 8) are studied. It is established that these molecules do indeed form hairpins with a helix of 6 AU base pairs and m cytidine residues in the loop. The thermodynamics of loop initiation are determined and found to be very different from expected values. The circular dichroism spectra of these molecules also give information about the conformation of the C residues in the loop.

B. Materials and Methods

1. Synthesis

The $(Ap)_5A(pC)_m(pU)_6$ oligomers ($A_6C_mU_6$) were synthesized enzymatically in three steps. First, poly A was hydrolysed, the terminal phosphates were removed with alkaline phosphatase, and the oligomers were separated to obtain A_6; second, C residues were added to the 3' end of A_6 with primer dependent polynucleotide phosphorylase and the different A_6C_m oligomers separated; and third, U residues were added to each AC block copolymer and separated again. The conditions for the first two steps were identical to those described by Martin, Uhlenbeck, and Doty (1971) for the synthesis of A_6U_n, except that CDP was used instead of UDP. The addition of U residues to A_6C_m was done under the following conditions: 0.54 mM A_6C_m, 17 mM H^3UDP (5.7

0 0 0 0 3 9 0 0 0 3 7

Ci/mole), 0.01 M $MgCl_2$, 0.6 M NaCl, 0.2 M Glycine buffer pH 9.3 and 0.3 mg/ml primer dependent polynucleotide phosphorylase from M. luteus. The reaction was run at 37°C for 2 hours.

Reaction products were separated on a 50 ml TEAE-cellulose column at pH 3.5 in 7 M urea. The reaction mixture was diluted with 7 M urea to an ionic strength of 0.02 M and applied to the column at neutral pH. The unreacted UDP was usually eluted off the column with 0.05 M NaCl, 0.01 M tris in 7 M urea at pH 8. Then the column was equilibrated with 0.01 M sodium formate, 7 M urea pH 3.5 and the oligomers were eluted with a gradient (900 ml total volume) from 0 to 0.275 M NaCl in 0.01 sodium formate, 7 M urea, pH 3.5. The elution profile for $A_6C_5U_N$ is shown in Figure 1. Fractions were pooled and desalted as described in Martin, et al. (1971).

Identification of the reaction products was accomplished by two methods. First, the specific activity of each peak could be compared with the fraction of U residues in the molecule and second, early peaks could be hydrolysed with RNase A, chromatographed on paper to separate U and Up, and the radioactive Up/U ratio determined thus giving the chainlength (Martin, et al., 1971).

2. Measurements

Absorption-temperature profiles were measured on

Figure 1.

TEAE-cellulose column elution profile of $A_6C_5U_N$ oligomers. The large first peak contains N = 0, 1. 2 as well as some of the UDP. The succeeding peaks are the $A_6C_5U_N$ series starting with N = 3. (See text for chromatographic details.)

0 0 0 0 3 9 0 0 7 3 8

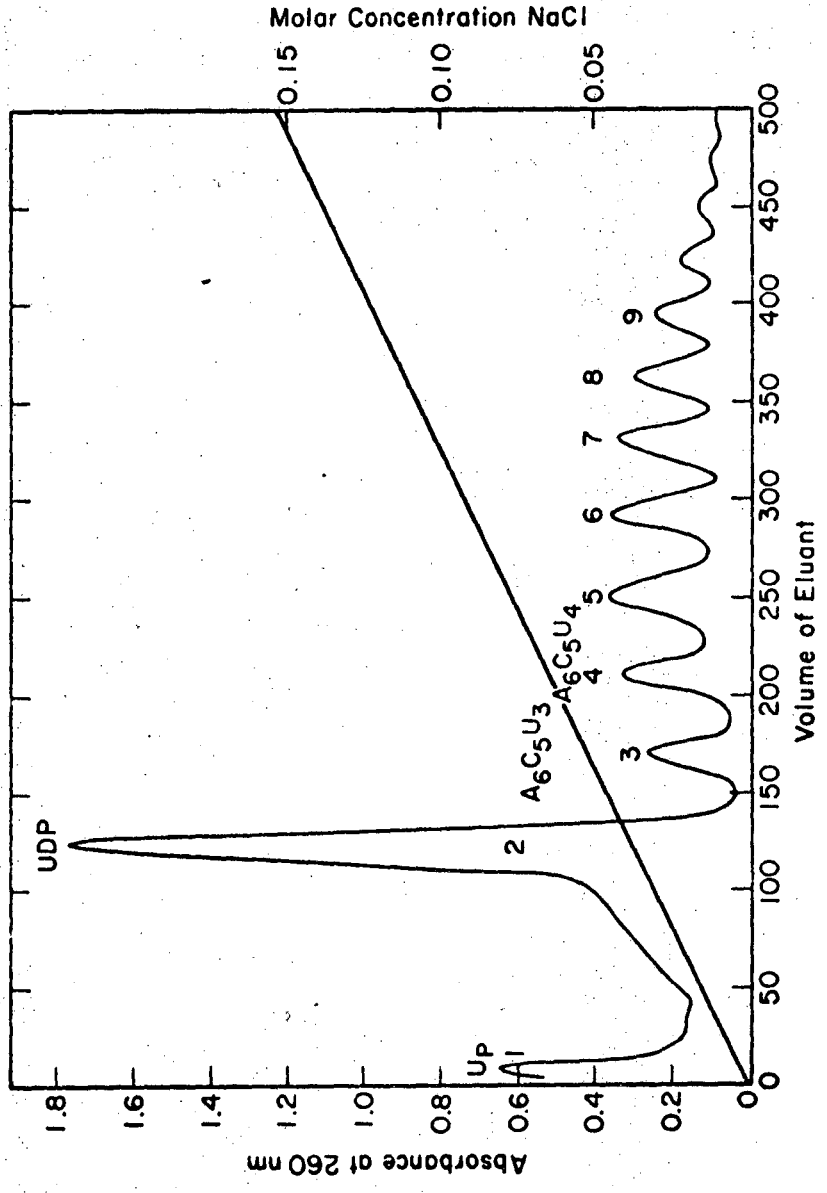


Figure 1

an automatic recording spectrophotometer (Gilford Instruments) as described in Martin, et al. (1971). The solvent consisted of 1 M NaCl, 0.01 M sodium phosphate buffer, and 10^{-4} M sodium EDTA, adjusted to pH 7.

Ultraviolet absorption spectra were obtained with a Cary 14 spectrophotometer equipped with a thermostatted circulating block.

Circular dichroism spectra were taken with a Cary 60 spectropolarimeter with a Model 6001 accessory. A PDP 8/S computer was used for on-line data acquisition and reduction (Tinoco and Cantor, 1970). The temperature was controlled by a thermostatted thermoelectric cooler (Allen, Gray, Roberts, and Tinoco, 1972). The spectra are presented as either circular dichroism ($\varepsilon_L - \varepsilon_R$) or ellipticity per mole of base.

Extinction coefficients were obtained by alkaline hydrolysis of the oligomer. The values obtained for ε_{260} at 47°C of $A_6C_mU_6$ are 9.3, 9.3, 9.1, 9.0 ($\times 10^3$) $\ell mole^{-1}cm^{-1}$ for m = 4,5,6 and 8 respectively.

3. Treatment of Melting Data

The absorbance-temperature profiles of the loop to coil transitions of the $A_6C_mU_6$ oligomers represent a sum of (1) the change in absorbance of the A-U double helix melting; (2) the change in absorbance of the C residues as the loop becomes a single strand; and (3) the change

0 0 0 0 3 9 0 0 7 3 9

in absorbance of the single strand with temperature. It is therefore not possible to evaluate the absorbance-temperature profile as a helix-coil transition unless certain assumptions and corrections are made. Figure 2 summarizes the conversion of a melting curve to a plot of fraction helix versus temperature. The melting curve, A(T), can be corrected for single stranded unstacking by using the absorbance change accompanying this unstacking, $A_S(T)$, as the upper bound of the melting curve (Martin, et al., 1971 and Figure 2a). $A_S(T)$ can be measured by melting at very low salt concentrations, where no helix is formed or it can be calculated from the unstacking of the component residues.

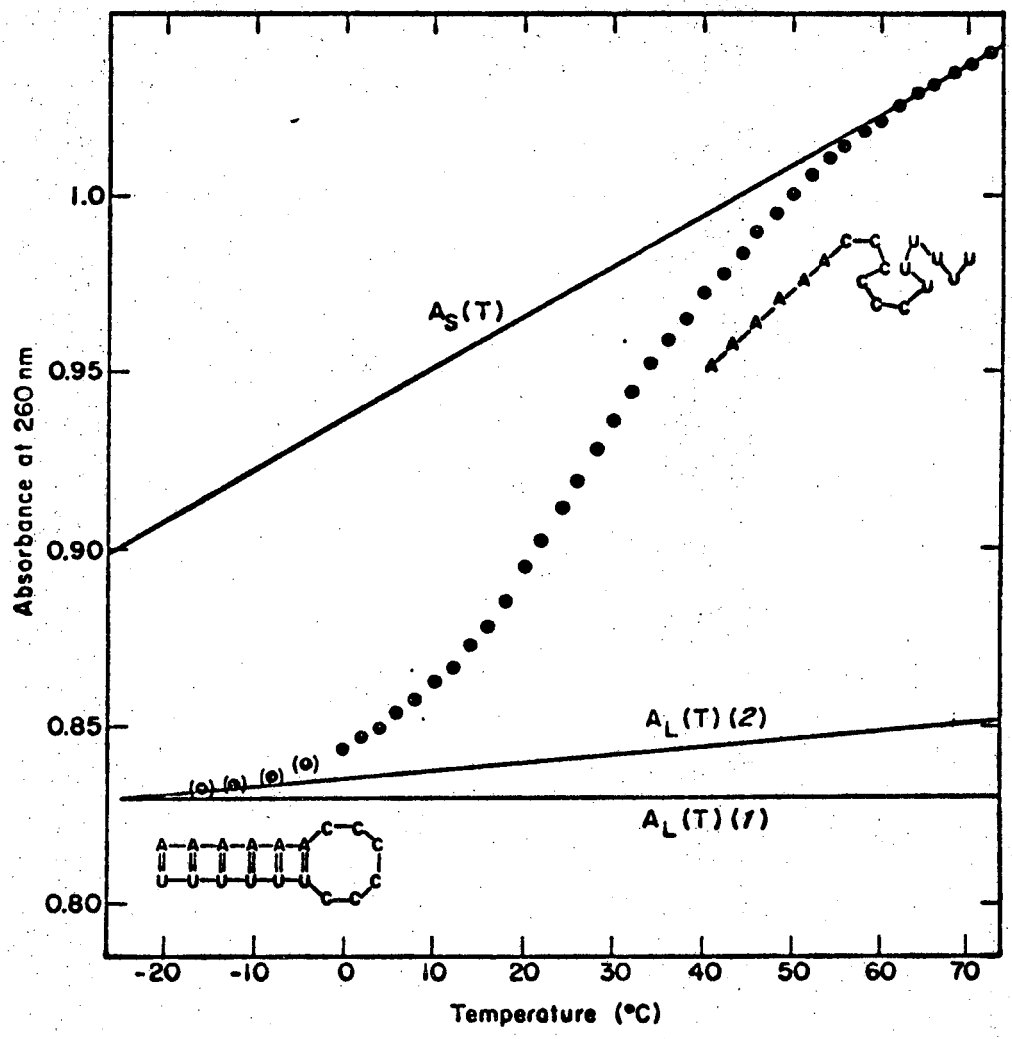
The absorbance of the fully formed loop, $A_L(T)$, is more difficult to obtain. In all the loops studied, the transition was quite broad and occurred at such low temperatures that at 0°C the loop was not fully formed. Thus, $A_L(T)$ must be obtained by extrapolation. Furthermore, since the conformation of the C residues within a closed loop may change with temperature, $A_L(T)$ might change with temperature. In order to obtain a reliable estimate of $A_L(T)$, the slope of A(T) is plotted as a function of temperature and extrapolated to low temperature. One can extrapolate to zero slope and choose $A_L(T)$ as independent of temperature and equal to the extrapolated absorbance. This is Method 1 shown in Figure (2). Alternately one can assume that the C's in

Figure 2.

(a) Absorbance-temperature profile of $A_6C_6U_6$ in 1 M NaCl, 10 mM sodium phosphate pH 7 normalized to 1.0 at 50°C and an average of five different concentrations. $A_S(T)$ is the absorbance of the oligomer as a single strand. $A_D(T)$ is the absorbance of the fully formed loop extrapolated to 0 slope in Method 1 and to a slope corresponding to oligo C unstacking in Method 2.

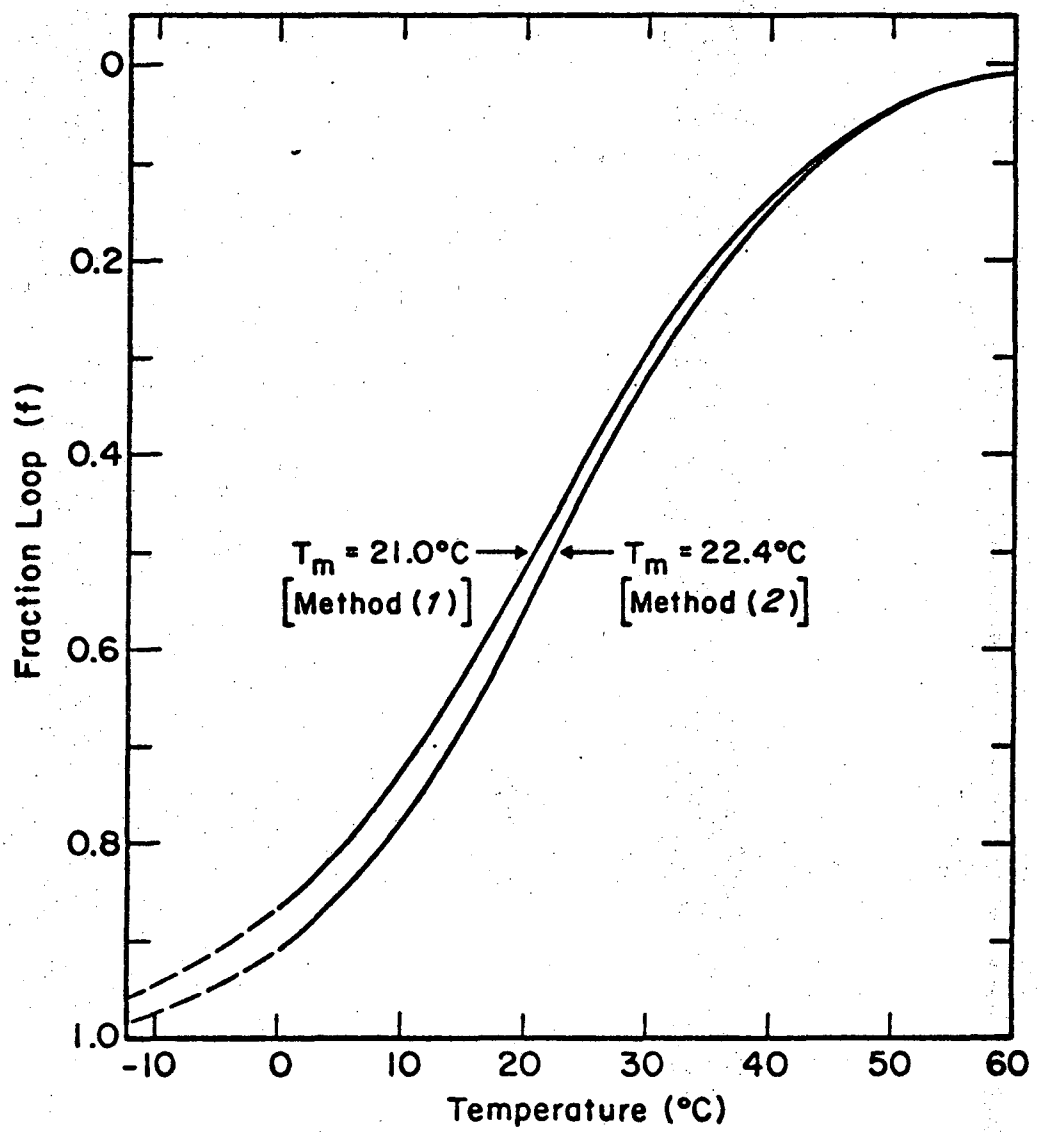
(b) Fraction of loop versus temperature for $A_6C_6U_6$ calculated from (a) by Methods 1 and 2.

0 0 0 0 3 9 0 0 7 4 0



XBL 723-458

Figure 2a



XBL 723-459

Figure 2b

0 0 0 0 3 9 0 0 9 4 1

the loop change their absorbance with temperature as much as single stranded oligo C does. In this case (Method 2) $A_L(T)$ is extrapolated to a final slope of 0.2%/10°C. For both methods, the fraction of loop, f, is then calculated as:

$$f = \frac{A_S(T) - A(T)}{A_S(T) - A_L(T)}$$

In Figure 2b, f versus T is given for the two alternative methods of estimating $A_L(T)$. Only a few degrees difference in the T_m (f = 0.5) and a small change in the slope is observed. We will base the results on Method 1, a constant value of $A_L(T)$, for simplicity.

C. Thermodynamic Results

1. Single Strand to Complete Loop

We want to study the thermodynamics of formation of intramolecular hairpin loops. The complete reaction can be written as:

$$A_6C_mU_6 \text{ (single strand)} \rightleftharpoons A_6C_mU_6 \text{ (loop)}$$

with equilibrium constant equal to f/(1-f) where f = loop fraction. The thermodynamic parameters which characterize the reaction are: T_m, the temperature where f = 0.5 (the equilibrium constant = 1); ΔH° (total), the standard enthalpy of the reaction; ΔS° (total), the standard entropy of the reaction. The value of ΔH° (total)

can be derived from the van't Hoff equation.

$$\Delta H^\circ \text{ (total)} = 4 R T_m^2 (\partial f/\partial T)_{T_m} \qquad (1)$$

This form is used because the value of the slope $(\partial f/\partial T)$ is most accurate at T_m where it is least affected by assumptions about extrapolated low and high temperature values for absorbance. This relation will underestimate the value of the total enthalpy if the reaction is not a two state equilibrium as shown. The standard free energy, ΔG°, equals 0 at T_m, so the standard entropy is given by:

$$\Delta S^\circ(\text{total}) = \Delta H^\circ(\text{total})/T_m \qquad (2)$$

Since both the loop → single strand transition and the unstacking of single strands are intramolecular reactions, the absorbance versus temperature profiles of the loops should be independent of concentration. Thus, for each $A_6C_mU_6$ (m = 4,5,6, and 8) at least four absorbance versus T curves were obtained over a twenty-five fold range of oligomer concentration. In each case, the normalized absorbance curves were independent of concentration to within experimental error. We concluded that only intramolecular loops formed and we averaged the data for the different concentrations. Further evidence for the formation of double stranded intramolecular helices came from the ultraviolet spectra of the loops. The spectral changes on melting of $A_6C_mU_6$ were consistent with double strand melting, while for oligomers with more than six uridines, the spectral

0 0 0 0 3 9 0 0 7 4 2

change is characteristic of triple strand formation, suggesting multistrand aggregates (Martin, et al., 1971).

If the absorbance of the completely formed loop is assumed to be independent of temperature (Method 1), the f versus T curves thus obtained for the four oligomers are shown in Figure 3. The thermodynamic parameters obtained from these curves are given in Table 1.

2. Loop Initiation

To learn about the process of closing the loop by the formation of the first base pair, we must consider the single strand to loop reaction in more detail. The simplest (all or none) assumption is that the reaction takes place in one step, and that the equilibrium constant for the reaction of $A_N C_m U_N$ can be written as:

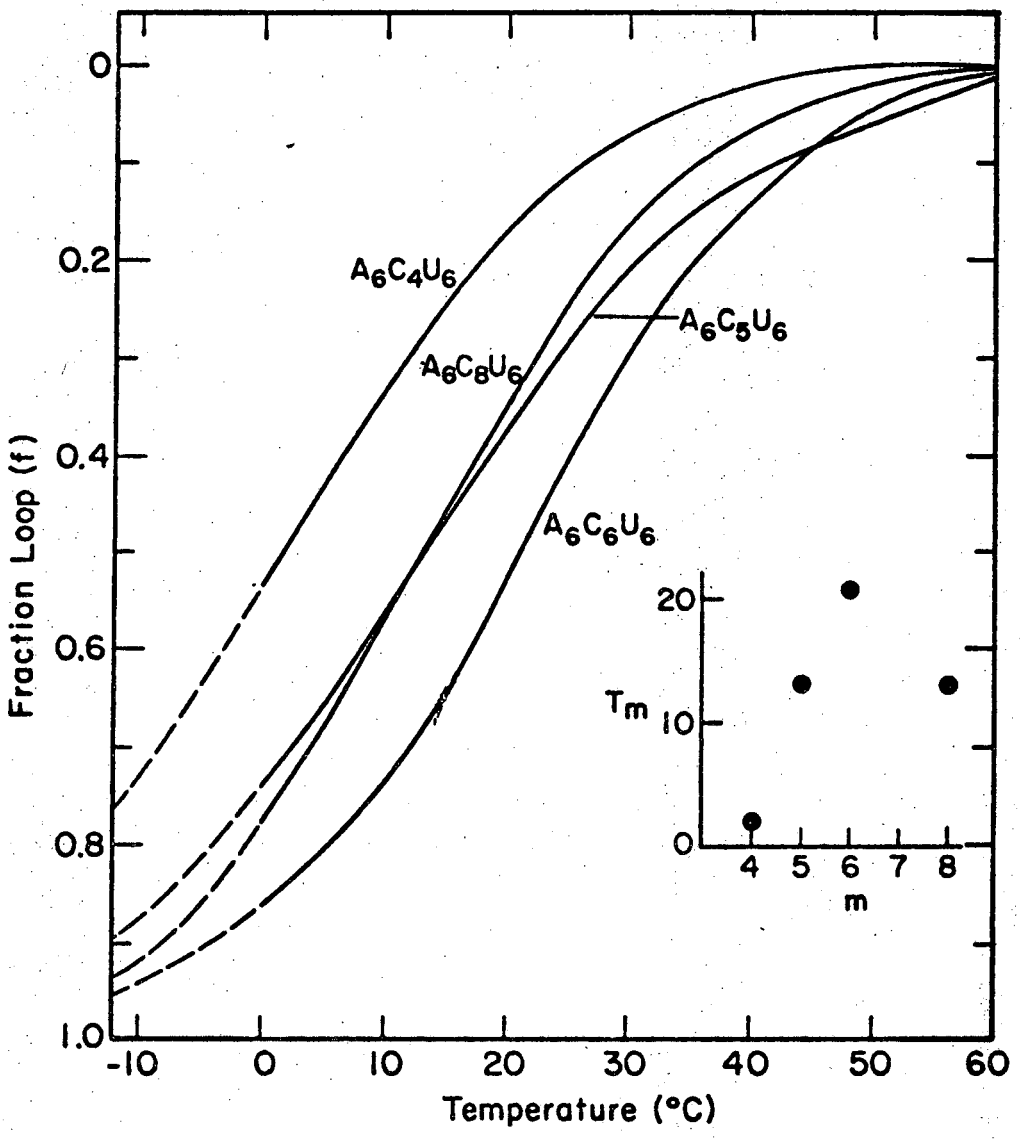
$$\frac{f}{1-f} = \gamma_m s^{N-1}$$

γ_m is the equilibrium constant for loop initiation for a loop of m unbonded bases. s is standard notation for the equilibrium constant for adding each successive base pair to an initiated helix. Both s and γ_m are expected to be temperature dependent. The appropriate equations follow immediately.

Figure 3.

Loop fraction versus temperature for $A_6C_mU_6$ m = 4, 5, 6 and 8 in 1 M NaCl, 10 mM sodium phosphate pH 7 calculated by Method 1. Each curve is an average of at least four different oligomer concentrations. The dotted lines indicate extrapolation of low temperature data.

0 0 0 0 3 9 0 0 4 3



XBL 723-451

Figure 3

Table 1

Experimental Thermodynamic Parameters for the Reaction

$A_6C_mU_6$ (Single Strand) to $A_6C_mU_6$ (Loop)

Loop	T_m (°C)	ΔH° (total) (kcal/mole)	ΔS° (total) (eu)	%h[a]
$A_6C_4U_6$	2	-12	-44	17
$A_6C_5U_6$	13.4	-13	-45	19
$A_6C_6U_6$	21.0	-16	-54	17
$A_6C_8U_6$	13.4	-14	-51	19

[a]The percent hyperchromicity is calculated at 50°C as $\%h = 100 \dfrac{A_S(50°C) - A_L(-10°C)}{A_S(50°C)}$

0 0 0 0 3 9 0 0 7 4 4

$$s(T) = e^{-(\Delta\bar{H}^\circ/R)\left(\frac{1}{T} - \frac{1}{T_m^\infty}\right)}$$

$$\gamma_m(T_2) = \gamma_m(T_1)\, e^{-(\Delta H_i^\circ/R)\left(\frac{1}{T_2} - \frac{1}{T_1}\right)} \qquad (3)$$

In the equation for $s(T)$, T_m^∞ is defined as the temperature where $s = 1$. For the A-U base pairs in $A_N C_m U_N$ this is equal to the T_m for poly A:poly U at the same salt concentration. The standard enthalpies which appear in the equations are: ΔH_i°, enthalpy of loop initiation, and $\Delta\bar{H}^\circ$, enthalpy (per base pair) of helix propagation. For simplicity the enthalpies are chosen independent of temperature, although in general they should also vary with temperature. Applying the van't Hoff equation, that is, taking the derivative of the equilibrium constant with respect to temperature, we obtain

$$\frac{1}{T_m} = \frac{1}{T_m^\infty} + \frac{R}{(N-1)\Delta H^\circ} \ln \gamma_m(T_m) \qquad (4)$$

for the relation between melting point for the loop, T_m, and thermodynamic parameters. This equation differs from that given by Kallenbach (1968) by the appearance of (N-1) instead of N. He chooses the equilibrium constant for loop initiation to be proportional to s; this leads to $(N)\Delta H^\circ$. We follow Scheffler, et al. (1970) in the definition of γ_m; they discuss its advantage. T_m^∞ and $\Delta\bar{H}^\circ$ can be obtained directly from experiments on

poly A:poly U. $\Delta\bar{H}°$ can also be calculated from thermodynamic data on the melting of self complementary oligomers of the type A_NU_N. In 1 M NaCl we use: T_m^∞ = 78°C and $\Delta\bar{H}°$ = -8.0 kcal/mole (Tinoco, et al., 1971). When these values and the measured T_m for each oligomer are substituted into Equation (4), one obtains γ_m values for the four loops at their T_m. The enthalpy of initiation, ΔH_i°, for these oligomers can be obtained by subtracting (5).(-8.0) = -40 kcal/more from the enthalpy of the total reaction given in Table 1. ΔH_i° can then be used to calculate γ_m at 25°C from γ_m at T_m by Equation (3). The γ_m (25°C) and the corresponding free energies and entropies of initiation at 25°C are given in Table 2.

Data are also available for a loop with an altered stem (C. Cech, unpublished work). The loop formed by $A_5GC_5U_5$ is similar to the $A_6C_4U_6$ loop except that the base pair closing the four unbonded cytosines is a GC pair instead of an AU pair. The T_m of $A_5GC_5U_5$ is 19.7°C and its total $\Delta H°$ = -15 kcal/mole. The fact that this T_m is nearly 18°C higher than the T_m for $A_6C_4U_6$ is not surprising in view of the fact that the $A_5GC_5U_5$ stem contains 16.7% G-C base pairs. Using Equation (4) with T_m^∞ adjusted to 90°C to account for the higher GC content, we obtain $\gamma_m(25°C) = 0.5 \times 10^{-5}$, ΔG_i° = +7.3 kcal, ΔH_i° = +25 kcal and ΔS_i° = +59 eu. These values are very similar to those given in Table 2 for $A_6C_4U_6$.

Table 2

Calculated Thermodynamic Parameters at 25°C for Forming the First A-U Base Pair which Closes a Loop of C Residues

	All-or-None Model			
No. of bases in loop (m)	γ_m	ΔH_i° (kcal/mole)	ΔS_i° (eu)	ΔG_i° (kcal/mole)
4	0.7×10^{-5}	+28	+70	+7.1
5	1.6×10^{-5}	+27	+69	+6.6
6	2.6×10^{-5}	+24	+59	+6.3
8	1.4×10^{-5}	+26	+65	+6.6

	Model with Intermediate States		
γ_m	ΔH_i° (kcal/mole)	ΔS_i° (eu)	ΔG_i° (kcal/mole)
0.3×10^{-5}	+24	+55	+7.5
0.9×10^{-5}	+24	+57	+6.9
2.0×10^{-5}	+21	+49	+6.4
1.1×10^{-5}	+24	+58	+6.8

Up to this point, the thermodynamics of loop formation has been discussed in terms of a simple two-state model. Although it seems intuitively reasonable that this approximation is valid for loops containing only a few base pairs, it is important to test the validity of the approximation by computing the thermodynamics with a more rigorous model. Thus, a partition function which contains all the likely partially bonded intermediate states was derived; the contribution of each intermediate state to the absorbance was assessed; and the effective fraction of base-paired states was calculated (M. D. Levine, unpublished work). In this formulation, a statistical weight of 1 was assigned to the single strand. All possible loop sizes were considered for the partially bonded states. That is, for $A_N C_m U_N$ loops of m, m+1, m+2...m+2N-2 unbonded bases were counted as intermediate forms. In other words, not only could melting occur from either end of the helix, but staggered conformations with an uneven number of A's and U's in the loop were allowed. However, interior loops and "bulges" in the helix were excluded as they can easily be shown to be negligible for a helix which can at most contain 6 base pairs. The contribution of each intermediate state to the change in absorbance was assumed to be proportional to the number of base-pair stacking interactions. A nearest neighbor approximation of this sort has been found to be applicable

0 0 0 0 3 9 0 0 7 4 6

in many related contexts (Gray and Tinoco, 1970). The effective fraction of base-paired states could then be calculated as a function of temperature with γ_m and ΔH_i° as variable parameters. $\Delta\bar{H}^\circ$ was chosen equal to -8.0 kcal as before. Unlike in the all-or-none model, the parameters for each oligomer depend on all the other oligomers. For example, to calculate a melting curve for $A_6C_4U_6$ all possible loop sizes from the minimum loop of 4 unbonded bases to the maximum loop of 14 unbonded bases must be considered. That is, the values of $\gamma_4, \gamma_5 \cdots \gamma_{14}$ must all be specified. However, in this case, the most important variable is γ_4 corresponding to the actual number of residues in the fully formed loop. The other values of γ_m are used to calculate the intermediate, partially paired forms. Values of γ_m for $m > 8$ were calculated from

$$\frac{\gamma_{8+x}}{\gamma_8} = \frac{9}{9+x}^{3/2}$$

This is the chain length dependence for γ_m predicted for a single strand which has a Gaussian distribution of end-to-end distances (Jacobson and Stockmayer, 1950). The values for ΔH_i° for $m > 8$ were set equal to $\Delta H_i^\circ(m=8)$. γ_7 and $\Delta H_i^\circ(m=7)$ were chosen as the average of the corresponding values for m=6 and m=8. The values of the thermodynamic parameters at 25°C obtained from fitting the calculated values (by the complete partition function) of T_m and $(df/dT)_{T_m}$ to the measured values are given in

Table 2. The values are not too different from those obtained from the all-or-none model; however, the partition function shows that intermediate species are significant. This is particularly true for $A_6C_4U_6$ and $A_6C_5U_6$ where one base pair can break to produce a more favored loop.

D. Thermodynamic Discussion

The best way to discuss the results of the thermodynamic analysis of loop closure is in the context of what one should expect. For a flexible chain the probability of loop closure should increase as the chain length decreases. Ideally, this should continue until some minimum loop size is reached. Since model building indicates that the minimum loop size is 3 unbonded bases for polyribonucleotides, we expect γ_m to increase as m decreases, with γ_3 a maximum. The simplest theory of loop initiation based on a "random coil" single strand (Jacobson and Stockmayer, 1950) predicts in fact that γ_m should be proportional to $(m+1)^{-3/2}$. The only previously measured value for γ_m is that of $\gamma_4 = 30 \times 10^{-5}$ by Scheffler, et al. (1970) for deoxy(AT) oligomers in 0.5 M Na^+. They assumed that the value was independent of temperature, that ΔH_i° was zero. It does indeed seem reasonable for ΔH_i° to be zero or plus or minus 1 or 2 kcal, if the enthalpy of initiation is a result of

0 0 0 0 3 9 0 0 9 4 7

forming 2 hydrogen bonds (one base-pair) in aqueous solution in a completely flexible loop.

The experimental thermodynamic results are gathered in Tables 1 and 2. We immediately note that the measured T_m values and therefore the calculated γ_m values do not increase monotonically with decreasing numbers of C's in the loop. Although the T_m for the C_6 loop is greater than that of the C_8 loop, the T_m's of the C_4 and C_5 loops are less than the C_6 loop instead of more. Furthermore, the value of γ_4 calculated with the complete partition function is 1/100 the value of γ_4 for the deoxy(AT) oligomers. The γ_m values are also very temperature dependent; the ΔH_i° values are over 20 kcal. Finally, the free energies of initiation, which range from +7.5 to +6.4 kcal/mole are somewhat higher than expected. For example, when DeLisi and Crothers (1971) calculated ΔG_i° using detailed structural information and γ_4 from the deoxy(AT) oligomers, they obtained values of ΔG_i° which ranged from +3.4 to +3.9 kcal/mole for m=4 to m=8. The DeLisi and Crothers calculation was an attempt to improve on the estimates of ΔG_i° published by Tinoco et al. (1971) of +5.5 to +6.0 kcal/mole for m=4 to m=8. The critical point is, of course, that both calculations predict an increase in free energy of initiation of 0.5 kcal for a loop of 8 bases relative to that of four bases. The experimental result is a decrease of 0.7 kcal.

It is also rather remarkable that the ΔS_i° displayed

in Table 2 have such large positive values. The large entropy increase upon forming a more ordered structure apparently reflects a favorable change in water structure.

The discrepancy between the measured values and trends for ΔG°_i, ΔH°_i, and ΔS°_i and the theoretical estimates can be attributed partly to unstacking of cytosines when the single strand is bent into a loop. If the temperature dependence of the optical rotation of CpC is analyzed in terms of a two-state stacked or unstacked model, approximate values of ΔH°(unstacking = +7kcal and ΔS°(unstacking) = 25 eu are obtained (Davis and Tinoco, 1968). Therefore, independent of the size of the loop, the values of ΔH°_i and ΔS°_i in loop initiation are consistent with loss of about 3 CpC stacking interactions. Thus, either all the cytosines in the loop are partially unstacked or two cytosines are completely unstacked. Studies with molecular models of loops with 4 to 8 unbonded bases indicate that both these possibilities could occur. A model of a loop of seven residues in which two of the bases are completely unstacked and the other five are fully stacked has been carefully described by Fuller and Hodgson (1967). These results imply that the stability of a loop can depend on the base composition and sequence of the residues in the loop as well as those which close it. Uracil, which essentially does not stack, should be particularly important. RNA loops with only A, C, or G in them probably

show similar behavior to the C containing loops studied here. Loops which contain U (Particularly next to the helical region as in anticodon loops) may have smaller γ_m and ΔH_i° values near zero. These loops could be more reasonably compared with the theoretical estimates of ΔG_i° which ignore base-base interactions.

E. Circular Dichroism

The circular dichroism spectrum (CD) was measured for each of the $A_6C_mU_6$ as a function of temperature in 1 M NaCl at pH 7. The CD spectrum of each oligomer at 0°C is shown in Figure 4. However, even at 0°C these oligomers are not all in one conformational form. The fraction of helix varies from 0.54 for $A_6C_4U_6$ to 0.87 for $A_6C_6U_6$; this makes the spectrum quite complex and difficult to analyse. Thus, the 0°C spectrum is a mixture of the spectra of single strands, of the fully formed loop and, if intermediate states are important (as deduced from the partition function), of loops with incompletely formed helices. When temperature is increased, the proportions of these different forms change and a corresponding CD spectral change is seen. In Figure 5 the spectrum of $A_6C_6U_6$ is shown at four temperatures spanning the transition range. As was the case for the absorption-temperature studies, the differences between the low and high temperature CD spectra are due

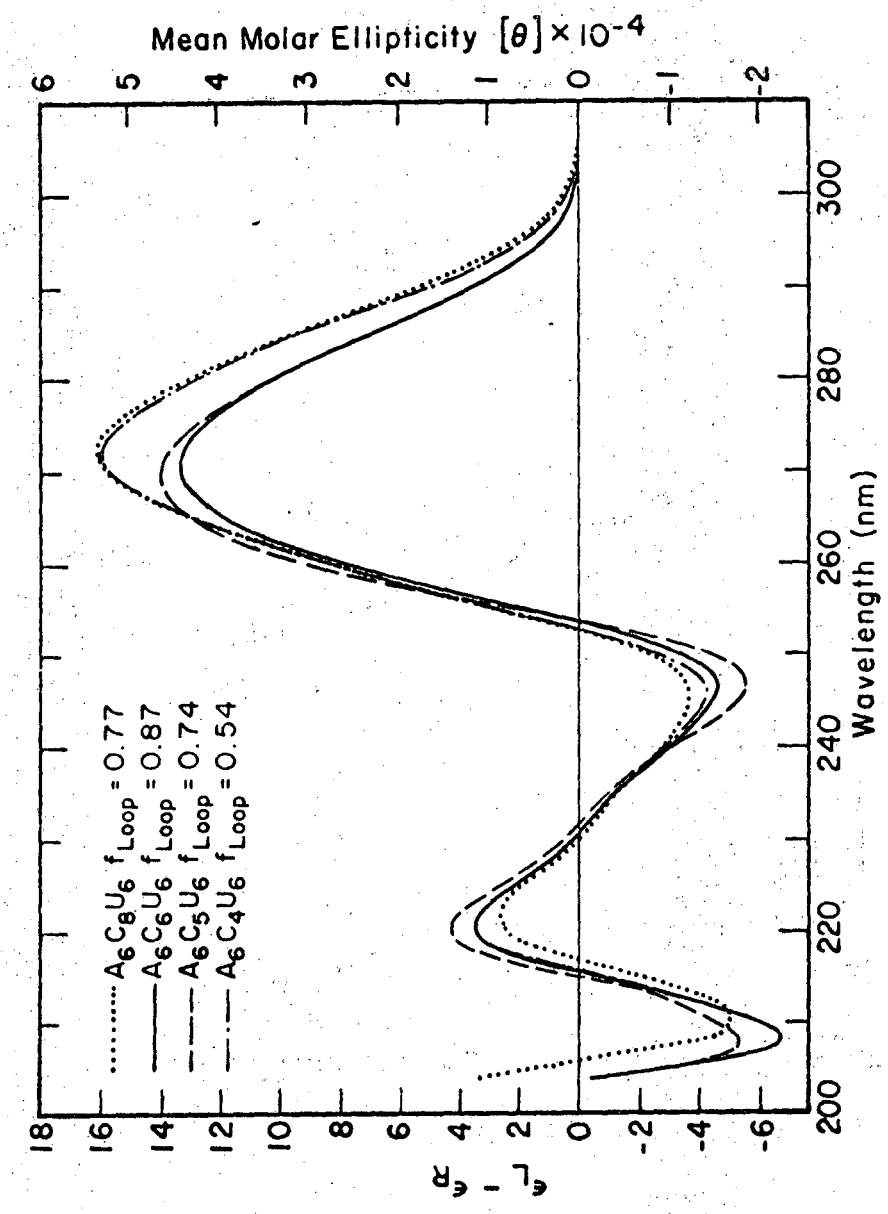
Figure 4.

CD spectra of $A_6C_mU_6$ m = 4,5,6 and 8 in 1 M NaCl, 10 mM sodium phosphate pH 7 and 0°C. The fraction of fully formed loop (f) for each oligomer is shown in the figure.

Figure 5.

CD spectra of $A_6C_6U_6$ at 0°, 20°, 40°, and 60°.

0 0 0 0 3 9 0 0 7 4 8



XBL 723-452

Figure 4

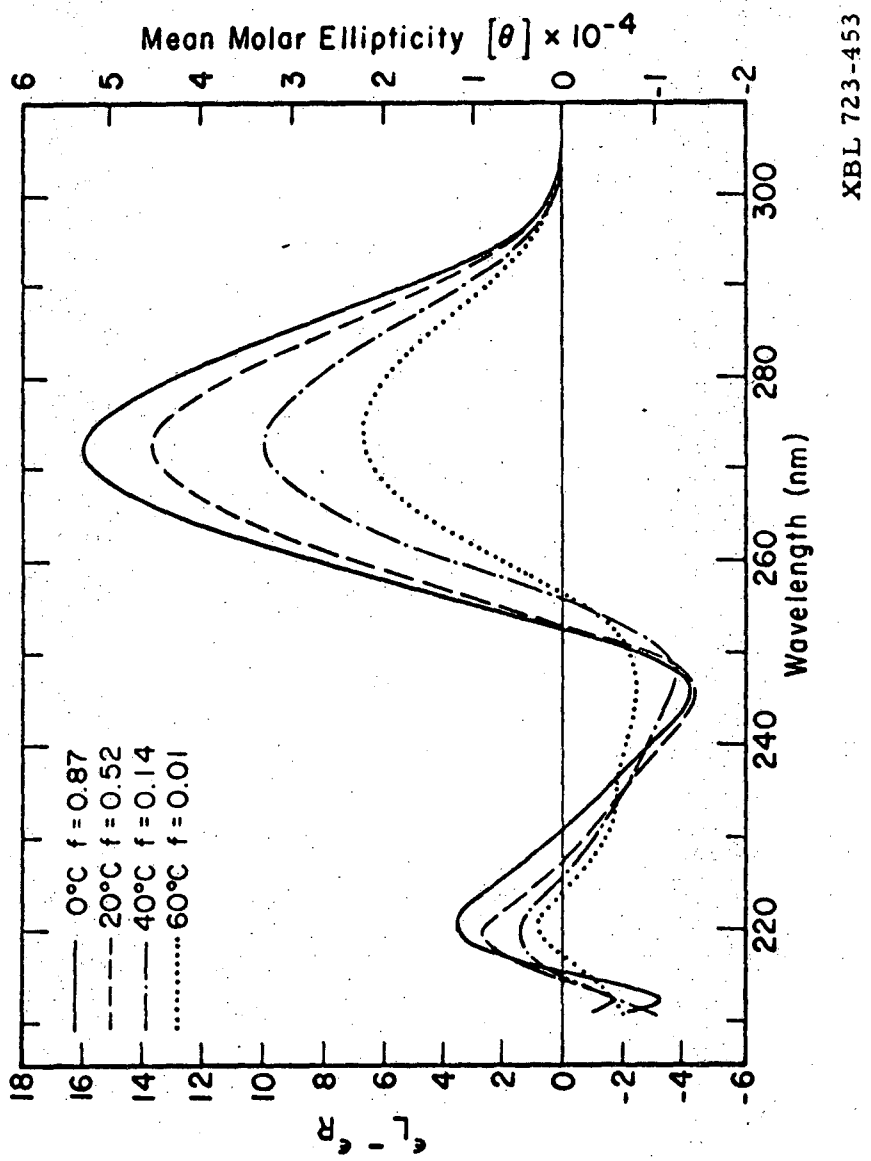

Mean Molar Ellipticity $[\theta] \times 10^{-4}$
6
5
4
3
2
1
0
-1
-2
300
280
260
240
220
Wavelength (nm)
0°C f = 0.87
20°C f = 0.52
40°C f = 0.14
60°C f = 0.01
18
16
14
12
10
8
6
4
2
0
-2
-4
-6
$\epsilon_L - \epsilon_R$
XBL 723-453

Figure 5

0 0 0 0 3 9 0 0 7 4 9

to (a) the change due to the AU pairs breaking, (b) the change due to unstacking of the single strands with temperature, and (c) the change accompanying the unstacking of the C residues when the loop is formed. The CD spectrum can be used to give qualitative information about the residues in the loop if we assume that the CD per base of the A_6:U_6 stem can be estimated from CD measurements of double stranded A-U base pairs.

In the simplest method, we assume that the CD spectrum of the cytosine residues in a loop can be obtained by subtracting the spectral contribution of the AU stem from the total CD spectrum of the loop. The AU contribution at different values of f was obtained from measured CD spectra of A_6U_6 at the same value of f. The double helix formed by A_6U_6 was chosen as a model for the stem region of the loop because, as an oligomer, it probably corresponds more closely to the conformation of the stem region than does the double stranded polymer poly A:poly U. Furthermore, A_6U_6 melts over a sufficiently wide range to obtain spectra at various values of f. Oligomer mixtures of A_N and U_N tend to form triple stranded aggregates in high salt concentrations and thus are not suitable models. We therefore measured the CD spectrum of a $A_6C_mU_6$ loop at various temperatures (corresponding to different f values) and subtracted the correct molar proportion of the A_6U_6 spectrum measured at the same value of f. In the $A_6C_6U_6$ case, the absorbance-

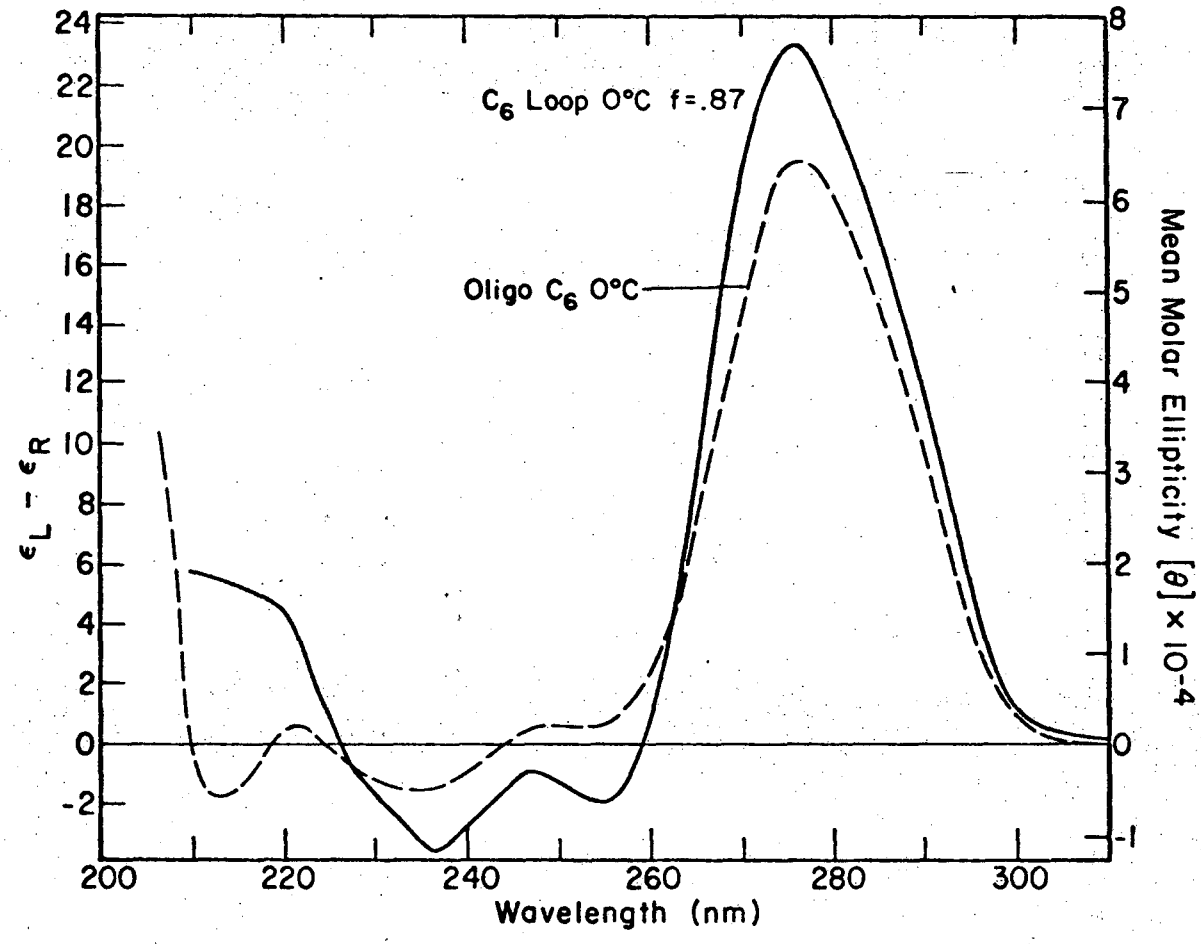
temperature profile is similar to that of A_6U_6 so the contribution of single strand stacking to the difference spectrum is quite small. The total contribution of the A-U residues is thus removed from the CD spectrum even though they are partially paired and partially single stranded. Single strand stacking appears more in the subtractions for the other loops since the temperature of the A_6U_6 spectrum must be higher for these less stable loops. This difficulty is most serious in the case of $A_6C_4U_6$ where, for example, an A_6U_6 spectrum at 25°C (f = 0.54) must be subtracted from the 0°C loop spectrum (f = 0.54). In Figure 6a and 6b, the CD spectrum of the cytosines in the C_6 loop of $A_6C_6U_6$ at 0°C are compared with the measured CD of the oligonucleotide $(Cp)_5C$. The comparison is made at 0°C and 40°C corresponding to f values of 0.87 and 0.14 for the loop. Three points are to be noted. First, there is excellent agreement of the general shape of the calculated spectrum of the C residues in the loop and the measured oligo C spectrum. This suggests that the loop C residues are in a conformation not radically different from what they are in single stranded oligo C. The second point is that the calculated CD spectra of the C residues in the loop change with temperature, implying that these residues are quite free to move while in the loop. Finally, the intensity of the calculated CD spectra is always somewhat greater than the corresponding

0 0 0 0 3 9 0 0 7 5 0

Figure 6.

Three times the CD spectrum of $A_6C_6U_6$ minus twice the CD spectrum of A_6U_6 at (a) f = 0.87 (0° for loop) and at (b) f = 0.14 (40° for loop) compared with the CD spectra of C_6 at 0° and 40°C.

Figure 6a



XBL 723-454

0 0 0 0 3 9 0 0 9 5 1

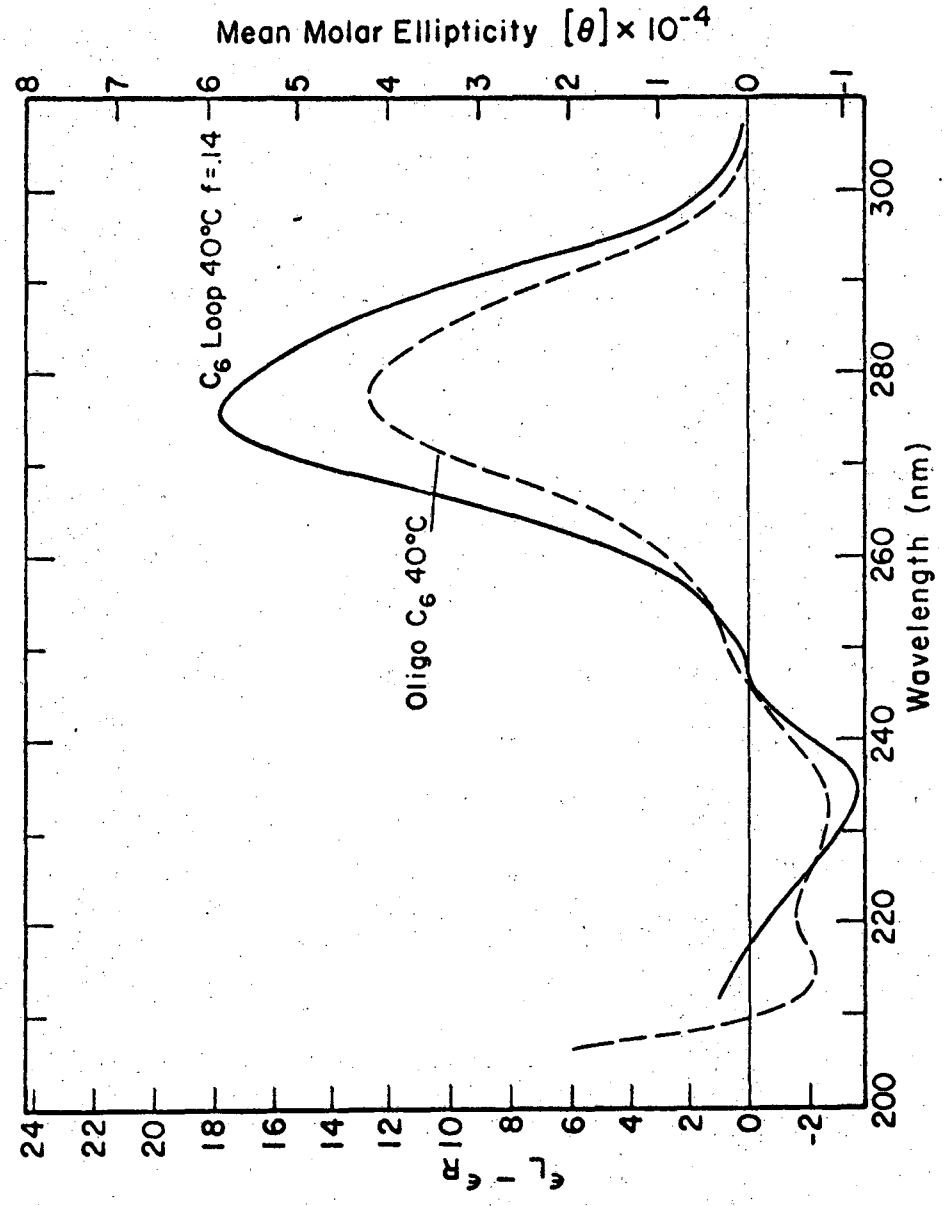


Figure 6b

measured oligo C at the same temperature. Although this may imply that C's in the loop have more structure than in the single strand, the many assumptions involved in the subtraction make quantitative conclusions suspect.

An attempt to improve the calculation of the CD of the loop residues was made by assuming nearest neighbor interactions for the AU interaction in the center of A_6U_6 and for the AC and CU interfaces between stem and loop. We correct for the fraction of single strand by subtracting an appropriate amount of the CD of the oligonucleotide C_m. This leads to a very similar shape and about a 5% increase in the maximum at 275 nm. Thus, no significant change from the simpler method is obtained.

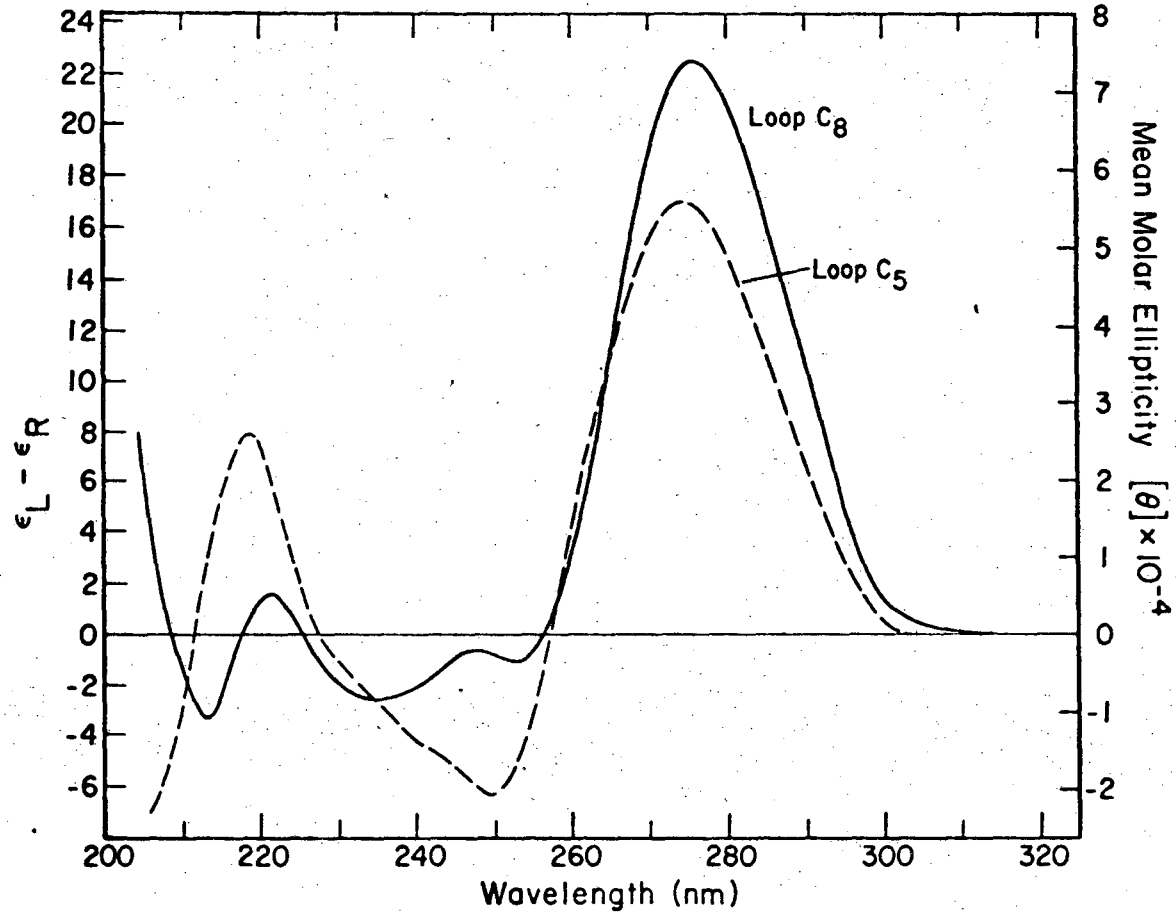
When the calculation of the CD spectrum of C residues in the loop is made for the other $A_6C_mU_6$ loops, an interesting observation is made. For m=8 the spectra are essentially the same as for m=6, but for m=5 the calculated spectra differ considerably in position and magnitude from the corresponding oligo C spectra. Figure 7 shows the contribution of the C residues in the loop for $A_6C_5U_6$ and $A_6C_8U_6$ at 0°C. These molecules both have about 75% helical structure at 0°C and are directly comparable. A similar comparison with the C_4 loop is not expected to be accurate since $A_6C_4U_6$ is only 54% helical at 0°C, leading to a large contribution from single strand stacking in the difference spectrum. The CD data imply that the C_5 loop has an altered conformation

0 0 0 0 3 9 0 8 7 5 2

Figure 7.

Calculated CD spectra of loop cytosines in $A_6C_5U_6$ at 0°C (f = 0.77). The appropriate amount of the CD of A_6U_6 at the same fraction helix is subtracted as in Figure 6.

Figure 7



XBL 723-456

0 0 0 0 3 9 0 0 9 5 3

from the C_6 and C_8 loops. This is consistent with the anomalously low stability found for C_4 and C_5 loops.

F. Conclusion

The RNA hairpin loops studied in this paper are of the same size range as those believed to occur in natural RNAs. Although the sequence of residues in the model loops could not be considered very natural, the information obtained from these molecules is probably quite applicable to their biological counterparts. The stability of the helical region closing the model loops was studied in detail. As was expected, the entropic effect of restraining complementary sequences to be close to one another did much to stabilize the helix. In our case, the T_m of two complementary strands with six AU base pairs has been estimated to be -13°C at 10^{-4}M strand concentration (Martin, et al., 1971), while the T_m of the loops are in the range of +10°C at all strand concentrations. The most striking result of the stability measurements was that the six mumbered loop was the most stable. Smaller loops were strained and larger loops showed the expected decrease in stability due to decreased probability of the A_6 and U_6 regions meeting. We note, of course, that two of the three loops in the well described cloverleaf model of tRNA always have seven residues. Unfortunately, we did not measure a loop

with seven bases, but we are confident that such a loop will have near maximal stability.

Little can be said about the conformation of the residues in the loop. The high free energy of initiation of the first helix base-pair suggests that either total unstacking of two loop residues or partial unstacking of all the residues must occur before the loop can form. This conclusion should be checked with loops which have U residues in them and therefore do not stack well, or with higher melting loops in which the loop bases unstack before the loop melts. The CD studies suggest that the average conformation of the C residues in the unstrained loops is not significantly different from the average single strand conformation of oligo C residues. The measurements are not sensitive enough to confirm the unstacking predicted by the thermodynamics. The average conformation of the C residues in the loop $A_6C_5U_6$ apparently is different from $A_6C_6U_6$, $A_6C_8U_6$, and oligo C since the calculated CD spectrum is considerably different. We expect similar behavior in the C_4 loop, but it is not currently possible to tell what this altered, presumably strained, conformation is.

These studies on RNA model loops are particularly useful for the prediction of RNA secondary structure from sequence (Tinoco, et al., 1971), since they provide the first information on the free energy of loop formation as a function of loop size. Future studies which

0 0 0 0 3 9 0 0 7 5 4

vary the composition and sequence of the residues in the loop and the helix which closes it should allow greatly improved predictive capacities.

This work was supported by grants from the National Institutes of Health (GM 10840 and GM 41393). We are very grateful to Mrs. C. Cech for allowing us to use her thermodynamic data on the $A_5GC_5U_5$ loop, and to Mr. M. Levine for calculations on the statistical mechanics of the intermediate states model. We are pleased to acknowledge helpful discussions with Mr. Levine, Mrs. Cech, Mr. E. Wickstrom, and Miss A. Blum.

REFERENCES

Allen, F.S., Gray, D.M., Roberts, G., and Tinoco, Jr. I. (1972). Biopolymers, in press.

Coutts, S.M. (1971). Biochim. Biophys. Acta, 232, 94.

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

DeLisi, C. and Crothers, D.M. (1971). Proc. Natl. Acad. Sci. U.S., 68, 2682.

Fuller, W. and Hodgson, A. (1967). Nature, 215, 817.

Gray, D.M. and Tinoco, Jr., I. (1970). Biopolymers, 9, 223.

Jacobson, H. and Stockmayer, W.H. (1950). J. Chem. Phys., 18, 1600.

Kallenbach, N.R. (1968). J. Mol. Biol., 37, 445.

Martin, R., Uhlenbeck, O.C., and Doty, P. (1971). J. Mol. Biol., 57, 201.

Römer, R., Riesner, P. and Maass, G., Wintermeyer, W., Thiebe, R., Zachau, H.G. (1969). FEBS Letters, 5, 15.

Scheffler, I.E., Elson, E.L. and Baldwin, R.L. (1970). J. Mol. Biol., 48, 145.

Tinoco, Jr., I. and Cantor, C.R. (1970). Methods of Biochemical Analysis, 18, 81.

Tinoco, Jr., I., Uhlenbeck, O.C., and Levine, M.D. (1971). Nature, 230, 362.

0 0 0 0 3 9 0 0 5 5

APPENDIX 1:

COMPUTER PROGRAMS

A. Program LISTIN

The purpose of this program is to make it extremely easy for an inexperienced programmer to set up an input deck for any program. The instructions for all the programs to follow use LISTIN to display sample input decks.

The program reads the contents of up to 500 cards and prints them out, each card image starting 50 spaces from the left margin to leave space for comments. At the top of the first page the columns 1 through 80 are indicated, and again at the end. By placing a ruler at the left of the number $\frac{4}{7}$ (column 47) at the top and bottom of the printout we see what goes in column 47 for any card. The number of cards must precede the cards to be displayed in a given set (I5 format), followed by the cards themselves, then other sets of data preceded by their count. A blank card terminates the program. The statements "PRINT 910" and "910 FORMAT(6H1*OVF*) turn off the automatic line skip for the printer at the beginning and end of a page. The format will probably have to be changed for use on a system other than the Lawrence Berkeley CDC computers.

```
      PROGRAM LISTIN(INPUT,OUTPUT)
      DIMENSION A(500,10)
   10 READ 900, NCARDS
       IF(NCARDS .EQ. -0.) STOP
      READ 901, ((A(I,J), J=1,10), I=1,NCARDS)
      PRINT 910
C
C THIS STATEMENT TURNS OFF TURNS OFF THE AUTOMATIC PRINTER LINE SKIP AT
C  THE BEGINNING AND END OF A PAGE.  I THINK THE FORMAT (910) WILL HAVE
C  BE CHANGED IF PROGRAM USED ON OTHER THAN THE HILL (LBL) COMPUTER.
C
      PRINT 902
      PRINT 903
      PRINT 905
      PRINT 904, ((A(I,J), J=1,10), I=1,NCARDS)
      PRINT 905
      PRINT 902
      PRINT 903
      GO TO 10
  900 FORMAT(I5)
  901 FORMAT(10A8)
  902 FORMAT (51X,* 12345678911111111112222222222333333333344444444445555
     1555555566666666677777777778*)
  903 FORMAT(60X,* 01234567890123456789C123456789012345678901234567890123456789012
     1345678901234567890*)
  904 FORMAT(52X,10A8)
  905 FORMAT(7)
  910 FORMAT(6H1*OVF*)
      END
```

LISTIN sample input:

```
12345678911111111112222222222333333333344444444445555555555666666666677777777778
        01234567890123456789012345678901234567890123456789012345678901234567890
```

```
    7
 THE PREVIOUS CARD SAYS THERE ARE 7 CARDS IN THIS SET.
THIS IS INPUT BEING OUTPUT.
    WHAT LISTIN READS, IT PRINTS.
A BLANK CARD TERMINATES READING NEW SETS OF DATA.  THERE IS ONLY ONE SET OF DATA
IN THIS EXAMPLE SO A BLANK FOLLOWS THIS, BELIEVE IT OR NOT.
```

```
12345678911111111112222222222333333333344444444445555555555666666666677777777778
        01234567890123456789012345678901234567890123456789012345678901234567890
```

B. Program SUPSPC

This program is used to read the paper tapes generated by the PDP 8/3 and convert them to cards. SUPSPC was written by Dr. M.S. Itzkowitz and has been previously described by Blum.[1] The input simply consists of a reel of paper tape in the proper format.[2] The output is a set of spectra punched in a 10F8.4 format. Immediately preceding the spectrum is the control block card containing the spectrum ID, the wavelength maximum, minimum (both in nm) interval between points (in Å), number of Å averaged per point, OD, and E. The format is: A8,2X,4(F8.3,2X),2(E13.6,2X). The card preceding the control card has all 80 columns filled with asterisks. Also, following each spectrum is another such asterisk card.

0 0 0 0 3 9 0 0 9 5 7

C. Program SMOOSP

This program takes the output from SUPSPC (raw data baselines and spectra) and plots them on a high speed line printer. The purpose of performing this procedure is to locate spurious data points. Occasionally the teletype punch or the card punch makes a mistake. These incorrect points are easily located by scanning the line printer plots and then repunching the bad card.

Most of the body of this program has to do with a generalized smoothing system for spectra. It is no longer used in generating smoothed CD spectra and the options for it are not described. All that is necessary to add to the output from SUPSPC is the "YES..." card (see sample input) and replacing the last asterisk card with a "STOP" card. The asterisk card immediately preceding the "CONTROL" card has been replaced in the sample input by a title card. This is not necessary but helps in keeping track of the identity of the spectra.

```
      PROGRAM SMOOSP(INPUT, OUTPUT, PUNCH)
C     PROGRAM APPLIES A 13 POINT SMOOTH (CUBIC) TO INPUT DATA.
C     INPUT-- IF IPLTR = YES, PROG PLOTS RAW DATA STARTING AT WAVELENGTH
C     XMAXR, IPLTS = YES, PROG PLOTS SMOOTHED DATA STARTING AT XMAXS,
C     YMAX,YINCR, AND ISY ARE PLOT PARAMETERS DESCRIBED IN SUBRCUTINE
C     PRNPLT AND GENERALLY CANNOT ALL BE LEFT BLANK. FOR PROPER INPUT
C     SEE READ 804, AND 804 FORMAT.  NEXT A ONE CARD IDENTIFIER FOR THE
C     SPECTRUM.  NEXT A CARD WITH THE CONTROL INFORMATION FROM SUPSPC,
C     SEE READ 801 AND 801 FORMAT.  NEXT, THE SPECTRUM IN 10F8.4 FORMAT.
C     NEXT A CARD WITH ALL 80 COLUMNS FILLED WITH ****S.  THEN NEW SETS
C     OF DATA AS ABOVE STARTING WITH THE IDENTIFIER.  PROGRAM IS
C     TERMINATED BY A CARD WITH 4HSTOP IN PLACE OF THE ASTERISK CARD.
      DIMENSION S(200), R(200), W(200), XIDENT(8)
  800 FORMAT(8A10)
  801 FORMAT(A8,2X, 4(F8.3,2X), 2(E13.6,2X))
  802 FORMAT(10F8.4)
  803 FORMAT(A6)
  804 FORMAT(2A3,4X,4F10.4,I5)
  806 FORMAT(// * RAW DATA FOR ID =* A8)
  807 FORMAT(// * SMOOTHED DATA FOR ID =* A8)
  808 FORMAT(// 3(15X,* LAMBDA*, 3X,* ELLIPTICITY*) //)
  810 FORMAT(1HI)
  811 FORMAT(80(1H*))
      M=0
      READ 804,  IPLTR, IPLTS, XMAXR, XMAXS, YMAX, YINCR, ISY
  101 READ 800 , (XIDENT(I), I=1,8)
      READ 801, ID, WMAX, WMIN, DELT, RPP, OD, E
      XINC = DELT/10.
      NPTS = (WMAX - WMIN)/XINC
      READ 802, (R(I), I=1,NPTS)
      READ 803, ASTER
      CALL SMOOTH(NPTS, R, S)
      IF(IPLTR .NE. 3HYES)  GO TO 121
      IF(M .EQ. NPTS) GO TO 115
      M = NPTS
      DO 110 I=1,NPTS
  110 W(I) = WMAX - FLOAT(I-1)*XINC
      NBEG = (WMAX - XMAXR)/XINC +1.
      NQ = NPTS - NBEG + 1
  115 CONTINUE
      PRINT 810
      PRINT 800, (XIDENT(I), I=1,8)
      PRINT 806, ID
      PRINT 808
      CALL PRNT(W,R,NPTS)
      PRINT 810
      PRINT 800, (XIDENT(I), I=1,8)
      PRINT 806, ID
      XINC = 1.65
      CALL PRNPLT(W(NBEG), R(NBEG),XMAXR,XINC, YMAX, YINCR, 0., ISY,NQ)
  120 CONTINUE
      IF(IPLTR .EQ. 3HYES)  GO TO 130
  121 CONTINUE
      IF (M .EQ. NPTS) GO TO 130
      M=NPTS
      DO 125 I=1,NPTS
  125 W(I) = WMAX - FLOAT(I-1)*XINC
  130 CONTINUE
```

```
      IF(IPLTS .NE. 3HYES)  GO TO 140
      NBEG = (WMAX - XMAXS)/XINC +1.
      NQ = NPTS - NBEG + 1
      PRINT 810
      PRINT 800, (XIDENT(I), I=1,8)
      PRINT 807, ID
      CALL PRNPLT(W(NBEG), S(NBEG),XMAXS,XINC, YMAX, YINCR, 0., ISY,NQ)
      PUNCH 800, (XIDENT(I), I=1,8)
      PUNCH801, ID, WMAX, WMIN, DELT, RPP, OD, E
      PUNCH 802, (S(I), I=1,NPTS)
      PUNCH 811
  140 CONTINUE
      IF (ASTER .EQ. 6HSTOP  ) STOP
      GO TO 101
      END


      SUBROUTINE SMOOTH(NPTS, R, S)
      DIMENSION S(200), R(200), T(13)
C      R = UNSMOOTHED DATA, S = SMOOTHED DATA, T = TEMP STORAGE
C      THE SIX LOWEST WAVELENGTH POINTS ARE NOT SMOOTHED AND ARE SET
C     EQUAL TO THE RAW DATA POINTS.  (ALSO THE SIX HIGHEST POINTS, BUT
C     USUALLY THEY ARE BEYOND THE RIGHT EDGE OF ANY PLOTS.
      DO 5 I=1,NPTS
    5 S(I) = R(I)
      N = NPTS - 12
      DO 10 I=2,13
      J= I-1
   10 T(I) = R(J)
      DO 200 I=1,N
      J = I+12
      DO 11 K= 1,12
      KK = K+1
   11 T(K) = T(KK)
      T(13) = R(J)
      SUM = 25.*T(7) + 24.*(T(6)+T(8)) + 21.*(T(5)+ T(9)) +16.*(T(4)+T(1
     10)) + 9.*(T(3)+T(11)) - 11*(T(1)+T(13))
      L = I+6
      S(L) = SUM/143.
  200 CONTINUE
      RETURN
      END


      SUBROUTINE PRNT(W,Y,NQ)
      DIMENSION W(200), Y(200)
  820 FORMAT (3(13X, F6.1, 3X,F13.4))
      LQ = NQ + 1
      DO 10 I=LQ,200
      W(I) = 0.
   10 Y(I) = 0.
      II = NQ/3. +1.
      DO 20 I=1,II
      J=I+II
      K=J+II
   20 PRINT 820, W(I),Y(I),W(J), Y(J),  W(K), Y(K)
      RETURN
      END
```

```
      SUBROUTINE PRNPLT(X,Y,XMAX,XINCR,YMAX,YINCR,ISX,ISY,NPTS)
C   PRINTER PLOT ROUTINE        M.S.ITZKOWITZ     MAY,1967
C
C     PLOTS THE *NPTS* POINTS GIVEN BY *X(I),Y(I)* ON A 51 X 101 GRID
C     USING A TOTAL OF 56 LINES ON THE PRINTER
C       IF *ISX* OR *ISY* ARE NON-ZERO, THE CORRESPONDING MAXIMUM AND
C        INCREMENTAL STEP SIZE ARE COMPUTED
C       IF EITHER INCREMENTAL STEP SIZE IS ZERO, THE PROGRAM EXITS
C       NEITHER OF THE INPUT ARRAYS ARE DESTROYED.  IF SCALING IS DONE
C       THE CORRESPONDING NEW VALUES OF MAXIMUM AND STEP SIZE ARE RETURNED
C
      DIMENSION X(NPTS),Y(NPTS),IGRID(105),XAXIS(11)
C
      INTEGER BLANK,DOT,STAR,IGRID,PLUS
      DATA BLANK,DOT,STAR,PLUS / 1H ,1H.,1H*,1H+ /
C
901   FORMAT(14X,105A1)
902   FORMAT(1XE10.3,2X,1H+,105A1,1H+)
903   FORMAT(15X,103(1H-))
904   FORMAT(7X,11(F10.0),2H (,I4,5H PTS)  )
905   FORMAT(16X,11(1H+,9X))
9800  FORMAT(46H1SCALING ERROR IN PRNPLT, EXECUTION TERMINATED )
C
      IF(ISX.NE.0) CALL PLSCAL(X,XMAX,XINCR,NPTS,100)
      IF(ISY.NE.0) CALL PLSCAL(Y,YMAX,YINCR,NPTS,50)
      IF(XINCR.EQ.0..OR.YINCR.EQ.0.) GO TO 800
      YAXMIN=0.01*YINCR
      XAXMIN=0.01*XINCR
      IZERO=YMAX/YINCR+1.5
      JZERO=103.5-XMAX/XINCR
      IF(JZERO.GT.103.OR.JZERO.LT.4) JZERO=2
      PRINT 905
      PRINT 903
      DO 10 I=1,51
      IF ( I.NE.IZERO) GO TO 16
      DO 14 J=1,105
14    IGRID(J)=PLUS
      GO TO 15
16    DO 11 J=1,105
11    IGRID(J)=BLANK
15    IGRID(JZERO)=PLUS
      IGRID(104)=DOT
      IGRID(2)=DOT
      DO 12 K=1,NPTS
      ITEST =(YMAX-Y(K))/YINCR+1.5
      IF(ITEST .NE.I) GO TO 12
      J=103.5-(XMAX-X(K))/XINCR
      IF(J.GT.103)J=105
      IF(J.LT.3) J=1
      IGRID(J)=STAR
12    CONTINUE
      IF(MOD(I,10).EQ.1) GO TO 13
      PRINT 901,IGRID
      GO TO 10
13    YAXIS=YMAX-(I-1)*YINCR
      IF(ABS(YAXIS).LT.YAXMIN) YAXIS=0.
      PRINT 902,YAXIS,(IGRID(J),J=1,105)
10    CONTINUE

      PRINT 903
      PRINT 905
      DO 20 M=1,11
      XAXIS(M)=XMAX-XINCR*(FLOAT(I-M))*10.0
      IF(ABS(XAXIS(M)).LT.XAXMIN)XAXIS(M)=0.
20    CONTINUE
      PRINT 904,XAXIS,NPTS
      RETURN
800   PRINT 9800
      CALL EXIT
      END
```

```
      SUBROUTINE PLSCAL(V,VMAX,VINCR,NPTS,NDIVIS)
C
C     SCALING PROGRAM FOR USE WITH PRNPLT    M.S.ITZKOWITZ  MAY,1967
C   THIS VERSION ADJUSTS THE FULL SCALE TO 2.5,5.0, OR 10. TIMES 10**N
C   AND ADJUSTS THE MAXIMUM POINT TO AN INTEGER MULTIPLE OF 5*VINCR
C
      DIMENSION V(NPTS)
C
      VMIN=V(1)
      VMAX=V(1)
      DO 10 I=1,NPTS
      IF(V(I).LT.VMIN) VMIN=V(I)
      IF(V(I).GT.VMAX) VMAX=V(I)
      QRANGE=VMAX-VMIN
10    CONTINUE
      IF(QRANGE.EQ.0.) GO TO 8000
      QRANGE=0.4342944*ALOG(QRANGE)
      IF(QRANGE)20,20,30
30    IRANGE=QRANGE
      GO TO 40
20    IRANGE=-QRANGE
      IRANGE=-IRANGE-1
40    QRANGE=QRANGE-FLOAT(IRANGE)
      RANGE=10.**QRANGE
C
C   RANGE IS BETWEEN 1.0 AND 10.0
C
43    IF(RANGE.GT.2.5) GO TO 41
      RANGE=2.5
      GO TO 50
41    IF(RANGE.GT.5.0) GO TO 42
      RANGE=5.0
      GO TO 50
42    RANGE=10.0
50    TRANGE=RANGE*(10.**IRANGE)
C
C   TRANGE IS NOW 2.5,5.0, OR 10.0 TIMES A POWER OF TEN
C
      VINCR=TRANGE/FLOAT(NDIVIS)
      IF(VMAX)51,51,52
52    IMAX=VMAX/(5.0*VINCR)
      XMAX=5.0*VINCR*FLOAT(IMAX+1)
      GO TO 53
51    IMAX=-VMAX/(5.0*VINCR)
      XMAX=5.0*VINCR*FLOAT(-IMAX+1)
53    IF(VMIN.GT.XMAX-TRANGE) GO TO 100
      RANGE=RANGE*2.0
      IF(RANGE-10.) 43,43,54
54    RANGE=RANGE/10.
      IRANGE=IRANGE+1
      GO TO 43
100   VMAX=XMAX
      VMIN=XMAX-TRANGE
      RETURN
8000  PRINT 9800
9800  FORMAT(45H1PLSCAL CALLED TO SCALE ARRAY WITH ZERO RANGE)
      CALL EXIT
      END
```

SMOOSP Input

```
1234567891111111111222222222233333333334444444444555555555566666666667777777778
01234567890123456789012345678901234567890123456789012345678901234567890123456790
```

	Plot card	YES 350. 1
	Identifier	BASELINE FOR A4U4 RUNS IN .126 MM CELL 7/20/72
Control card A8,2X,4(F8.3,2X),2(E13.6,2X) format, see SUPSPC description for more details.		72072808 350.000 190.000 10.000 8.000 1.004 80.3
		.1036 .1037 .1037 .1038 .1038 .1036 .1035 .1035 .1035 .1036
		.1037 .1037 .1037 .1037 .1038 .1037 .1038 .1038 .1038 .1038
		.1038 .1036 .1037 .1038 .1038 .1038 .1037 .1038 .1038 .1039
	Spectral points --	.1041 .1039 .1038 .1041 .1042 .1039 .1042 .1042 .1042 .1045
		.1044 .1044 .1044 .1044 .1042 .1044 .1046 .1046 .1046 .1046
	Format(10F8.4)	.1047 .1045 .1046 .1045 .1048 .1048 .1050 .1048 .1049 .1049
		.1048 .1049 .1049 .1050 .1050 .1050 .1050 .1050 .1050 .1049
		.1050 .1050 .1052 .1053 .1050 .1050 .1052 .1052 .1053 .1052
		.1052 .1053 .1053 .1053 .1052 .1053 .1053 .1052 .1055 .1053
		.1055 .1055 .1055 .1056 .1055 .1053 .1056 .1055 .1057 .1058
		.1058 .1058 .1059 .1060 .1063 .1060 .1061 .1063 .1064 .1069
		.1069 .1073 .1073 .1080 .1078 .1083 .1086 .1089 .1089 .1091
		.1095 .1100 .1100 .1106 .1103 .1108 .1110 .1115 .1119 .1119
		.1117 .1122 .1125 .1124 .1119 .1127 .1125 .1133 .1135 .1136
		.1133 .1136 .1138 .1139 .1141 .1139 .1144 .1144 .1152 .1144
		.1131 .1132 .1149 .1150 .1125 .1050 .1071 .0779 .0702 .0809
	Asterisk card	**********
	next identifier	A4U4 IN .126 MM CELL -4 DEG
	new control	A4U4-4 350.000 190.000 10.000 8.000 1.004000E+00 1.026000E+02
		.1042 .1039 .1042 .1042 .1041 .1044 .1042 .1042 .1042 .1044
		.1044 .1042 .1042 .1044 .1042 .1044 .1042 .1045 .1044 .1044
	more spectral	.1044 .1044 .1042 .1044 .1044 .1045 .1046 .1046 .1047 .1047
	points	.1048 .1048 .1048 .1050 .1050 .1050 .1050 .1050 .1050 .1052
		.1052 .1052 .1053 .1056 .1056 .1058 .1059 .1063 .1060 .1061
		.1064 .1066 .1069 .1075 .1080 .1089 .1099 .1111 .1128 .1147
		.1167 .1197 .1229 .1265 .1298 .1327 .1361 .1385 .1412 .1432
		.1451 .1471 .1498 .1520 .1548 .1576 .1608 .1633 .1653 .1641
		.1693 .1703 .1716 .1724 .1720 .1720 .1703 .1686 .1666 .1641
		.1595 .1556 .1497 .1443 .1355 .1273 .1191 .1121 .1042 .0957
		.0956 .0929 .0915 .0909 .0914 .0933 .0959 .0981 .1006 .1023
		.1055 .1073 .1095 .1114 .1136 .1147 .1164 .1177 .1194 .1204
		.1211 .1227 .1246 .1259 .1278 .1293 .1309 .1317 .1315 .1302
		.1276 .1257 .1219 .1177 .1125 .1082 .1017 .0987 .0963 .0956
		.0956 .0941 .0955 .1012 .0995 .1064 .1081 .1124 .1151 .1183
		.1177 .1227 .1283 .1315 .1392 .1289 .1292 .1324 .1444 .1158
	&HSTOP replaces	STOP
	last asterisk	
	card	

```
1234567891111111111222222222233333333334444444444555555555566666666667777777778
01234567890123456789012345678901234567890123456789012345678901234567890123456790
```

0 0 0 0 3 9 0 0 7 6 0

D. Program GETNSM

This program calculates and smooths CD and performs a baseline correction by shifting the spectrum such that the average difference between spectrum and baseline is zero for points above 325 nm. A blank card can be substituted for a "star" card at any time.

"B" input (normal) is used for spectra punched as raw data with only one baseline. The OD and E punched for the baseline are the ones used for all spectra following. "B" input is used primarily for cases where the sample cuvette was infrequently changed, e.g., temperature dependence or multiple scans.

"DW" input is used most frequently when succeeding spectra have the same baseline but different OD or E, e.g., distilled water spectra at one temperature.

"B2" input is used to average two baselines, then proceed to calculate CD. It may be used in conjunction with "DW" input or normal input in which case the OD and E on the first baseline is used for the following spectra.

"R-B" input is used for spectra which have been punched as raw spectrum - baseline. Use of this option in such a case effects baseline corrections and/or correcting an inappropriate OD or E. The OD and E in the baseline are used eventually to calculate the CD. The OD and E in the spectrum must be identical to those used in the original generation of punched data by the PDP 8/S. The

"R-B" option cannot be used in conjunction with "B2". "DWRB" input is presently not working correctly and since the occasion to use it has not so far occurred it will probably remain that way. Its intention was to provide for handling data in both "R-B" and "DW" conditions and will work okay if no change in OD or E is desired. The option is activated by following the baseline with a card having "DWRB" punched in the first four columns. The rest is the same as for R-B input. A new baseline will have to be input for each spectrum (if OD or E change).

Unlisted routines required:

PRNT	(see SMOOSP listing)
PRNPLT	(see SMOOSP listing)
PLSCAL	(see SMOOSP listing)
SMOOTH	(see SMOOSP listing)

```
      PROGRAM GETNSM(INPUT,OUTPUT, PUNCH)
C
C PROGRAM GETNSM = CDGET + SMOOSP, 6/29/72  PHIL BOREK
C
      DIMENSION B(200) ,R(200),RNAME(8), BNAME(8), W(200)
      COMMON/ARRAYS/R,RNAME,W
      COMMON/VARS/ID,WMAX,WMIN,DELT,RPP,CD,E,IPLTS,XMAXS,YMAX,YINCR,
     1  ISY,NPTS,XINC,XINCS
      COMMON/A/ PRFILE(10C,8), IPFIL, PIDFIL(100)
      IPFIL = 1
      OLDCHK = 4H****
    1 READ 909,CHK
      IF(OLDCHK.EC.3HR-B .AND. CHK .EC. 4H****) CHK = OLDCHK
      IF(OLDCHK.EC.3HR-B .AND. CHK .EC. 4H    ) CHK = OLDCHK
      IF(OLDCHK.EC. 2HDW .AND. CHK .EQ. 4H****) CHK = OLDCHK
      IF(OLDCHK.EC. 2HDW .AND. CHK .EC. 4H    ) CHK = OLDCHK
      IF(OLDCHK .EQ. 4HDWRB .AND. CHK .EQ. 4H****) CHK  =   OLDCHK
      IF(OLDCHK .EQ. 4HDWRB .AND. CHK .EQ. 4H    ) CHK  =   OLDCHK
      IF(CHK .NE. 4HSTOP) GO TO 3
      IPFIL = IPFIL - 1
      PRINT 905
      PRINT 909
      DC 2 J=1,IPFIL
      PRINT 901, (PRFILE(J,I), I=1,8)
    2 PRINT 910,PIDFIL(J)
  909 FCRMAT(* THIS IS THE FILE OF PLOTTED SPECTRA AND THEIR 8 CHARACTER
     1 ID#S.*//)
  910 FCRMAT(* WHICH CARRIES THE ID,* ,1X, A8,/)
      STOP
    3 CONTINUE
      IF(CHK .EQ. 1HB) GO TO 4
      IF(CHK .EQ.2HB2) GC TO 4
      GO TO 10
    4 CONTINUE
      READ 804,IPLTS, XMAXS,YMAX,YINCR,ISY,XINCS
      READ 901,(BNAME(I), I=1,8)
      READ 902, ID,WMAX,WMIN,DELT,RPP,DD,E
      XINC = DELT/10.
      NPTS=(WMAX-WMIN)/XINC
      CC 5 I=1,NPTS
    5 W(I) = WMAX - FLOAT(I-1)*XINC
      READ 903, (B(I), I=1,NPTS)
      IF (CHK .NE. 2HB2) GO TO 9
      READ 900, DUM
      GC TO 10
    9 READ 900, CHK
   10 READ 901,(RNAME(I), I=1,8)
      READ 902, ID,WMAX,WMIN,DELT,RPP,DOD,DE
      IF(CHK .EQ.  2HDW) GO TO 101
      IF(CHK .EQ.4HDWRB) GC TO 101
      GO TO 11
  101 CONTINUE
      E = DE
       CD = COD
   11 CONTINUE
      XINC=DELT/10.
      NPTS=(WMAX-WMIN)/XINC
      IF(XINCS .EC. -0.) XINCS = XINC
```

```
      READ 903, (F(I), I=1,NPTS)
      OLDCHK = CHK
      IF(CHK.NE.4HB2  ) GO TO 20
      CALL BASAVG(R,B,NPTS)
      PRINT 905
      PRINT 901, (BNAME(I), I=1,8)
      PRINT 906,ID
      PRINT 908
      CALL PRNT(W,B,NPTS)
      GO TO 1
   20 CONTINUE
      IF (CHK .EQ. 3HR-B) GO TO 201
      IF (CHK .EQ.4HOWRB) GO TO 201
      GO TO 24
  201 CONTINUE
      DO 21 I=1,NPTS
   21 R(I) = R(I)*DOD/DE + B(I)
   24 CONTINUE
      NC = (WMAX - 325.)/XINC
      COR = 0.
      DO 25 I=1,NC
   25 COR = COR + R(I) - B(I)
      COR = COR/NC
      DO 30 I=1,NPTS
      W(I) = WMAX - FLOAT(I-1)*XINC
   30 R(I) = (R(I) - COR - B(I)) *E/OD
      COR = COR*E/OD
      PRINT 905
      PRINT 901, (RNAME(I), I=1,8)
      PRINT 907,ID, COR
      PRINT 908
      CALL PRNT(W,R,NPTS)
       CALL SMOOSP
      GO TO 1
  804 FORMAT(A3,7X,3F10.4,I5,F15.4)
  900 FORMAT(A4)
  901 FORMAT(8A10)
  902 FORMAT(A8,2X,4(F8.3,2X),2(E13.6,2X))
  903 FORMAT(10F8.4)
  904 FORMAT(80(1H*))
  905 FORMAT(1H1)
  906 FORMAT(//* BASELINE FOR ID=* A8)
  907 FORMAT(//* RAW DATA SPECTRUM FOR ID=* A8, * CCR =*,F5.3)
  908 FORMAT(// 3(15X,* LAMBDA*,3X,* ELLIPTICITY*)//)
      END
```

0 3 0 0 3 9 0 0 9 6 2


```
      SUBROUTINE EASAVG(R,B,NPTS)
      DIMENSION R(200),B(200)
      DO 10 I=1,NPTS
   10 B(I) = (B(I)+R(I))/2.
      RETURN
      END

      SUBROUTINE SMOOSP
      DIMENSION S(200), R(200), T(13)
      DIMENSION XIDENT(8)
      COMMON/ARRAYS/R,XIDENT,W(200)
      COMMON/VARS/ID,WMAX,WMIN,DELT,RPP,OD,E,IPLTS,XMAXS,YMAX,YINCR,
     1  ISY,NPTS,XINC,XINCS
      COMMON/A/ PRFILE(100,8), IPFIL,  IDFIL(100)
  800 FORMAT(8A10)
  801 FORMAT(A8,2X, 4(F8.3,2X), 2(E13.6,2X))
  802 FORMAT(10F8.4)
  807 FORMAT(// * SMOOTHED DATA FOR ID =* A8, * E =*,E15.8,* OD=*,E15.8)
  808 FORMAT(// 3(15X,* LAMBDA*, 3X,* ELLIPTICITY*) //)
  810 FORMAT(1H1)
  811 FORMAT(80(1H*))
      CALL SMOOTH(NPTS, R, S)
      IF(IPLTS .NE. 3HYES)  GO TO 140
      NBEG = (WMAX - XMAXS)/XINC +1.
      NQ = NPTS - NBEG + 1
      PRINT 810
      PRINT 800, (XIDENT(I), I=1,8)
      PRINT 807, ID,E,OD
      CALL PRNT(W,S,NPTS)
      PRINT 810
      PRINT 800, (XIDENT(I), I=1,8)
      PRINT 807, ID,E,OD
      CALL PRNPLT(W(NBEG), S(NBEG),XMAXS,XINCS,YMAX, YINCR, 0., ISY,NQ)
      PUNCH 800, (XIDENT(I), I=1,8)
      PUNCH801, ID, WMAX, WMIN, DELT, RPP, OD, E
      PUNCH 802, (S(I), I=1,NPTS)
      PUNCH 811
      DO 7 I=1,8
    7 PRFILE(IPFIL,I) = XIDENT(I)
      IDFIL(IPFIL) = ID
      IPFIL = IPFIL + 1
  140 CONTINUE
      RETURN
      END
```

```
                                                  12345678911111111112222222222333333333344444444445555555555666666666677777777778
Normal "B" input to GETNSM                                  01234567890123456789012345678901234567890123456789012345678901234567890

All baselines preceded with B or B2 card          B
Plot card (see below) Format(A3,7X,3F10.4,I5,F15.4) YES       310.       6.        .2          0       1.2
Title card                                        BASELINE FOR A4UAU4 RUNS IN 1 CM CELL IN TMTB 6/28/72
Control card (see SUPSPC description), OD and E   628728LE   350.000   205.000    10.000     8.000    .939          102.
    here are used for all following calculations,   .1089   .1089   .1088   .1088   .1088   .1086   .1086   .1086   .1085   .1086
    until new B card found.                         .1085   .1086   .1085   .1084   .1083   .1085   .1083   .1083   .1081   .1082
Baseline raw data Format(10F8.4)                    .1083   .1081   .1082   .1080   .1081   .1081   .1081   .1081   .1081   .1078
                                                    .1081   .1080   .1081   .1081   .1081   .1081   .1078   .1081   .1078   .1078
                                                    .1081   .1080   .1080   .1080   .1081   .1078   .1080   .1080   .1078   .1080
                                                    .1080   .1078   .1080   .1078   .1077   .1078   .1078   .1080   .1078   .1080
                                                    .1080   .1078   .1078   .1078   .1078   .1078   .1078   .1075   .1075   .1078
                                                    .1077   .1077   .1077   .1077   .1075   .1078   .1075   .1073   .1078   .1077
                                                    .1075   .1075   .1074   .1075   .1077   .1075   .1078   .1075   .1078   .1077
                                                    .1075   .1075   .1078   .1075   .1077   .1075   .1075   .1075   .1075   .1075
                                                    .1077   .1074   .1075   .1075   .1075   .1077   .1075   .1073   .1077   .1075
                                                    .1081   .1082   .1084   .1088   .1088   .1088   .1093   .1094   .1097   .1096
                                                    .1100   .1100   .1099   .1106   .1108   .1106   .1108   .1108   .1108   .1106
                                                    .1108   .1107   .1102   .1100   .1100   .1097   .1091   .1093   .1085   .1080
                                                    .1085   .1080   .1067   .1044   .1035
Asterisk (Star) card                              ****************************************************************************
Spectrum title                                    A4UAU4 IN TMTB, 1 CM CELL    -3.5 DEG
Control card                                      A4UAU4-4   350.000   205.000    10.000     8.000   9.390000E-01   1.020000E+02
                                                    .1089   .1088   .1088   .1086   .1088   .1086   .1086   .1085   .1084   .1085
                                                    .1086   .1086   .1084   .1084   .1084   .1085   .1083   .1083   .1082   .1082
Spectrum raw data                                   .1080   .1078   .1080   .1080   .1081   .1081   .1081   .1081   .1081   .1081
                                                    .1081   .1080   .1082   .1081   .1080   .1080   .1080   .1078   .1080   .1078
                                                    .1080   .1080   .1080   .1080   .1080   .1080   .1080   .1078   .1080   .1080
                                                    .1081   .1080   .1083   .1084   .1083   .1086   .1092   .1097   .1106   .1111
                                                    .1124   .1138   .1155   .1174   .1193   .1213   .1233   .1250   .1263   .1277
                                                    .1294   .1306   .1322   .1338   .1355   .1376   .1398   .1415   .1432   .1443
                                                    .1458   .1471   .1483   .1490   .1497   .1499   .1493   .1490   .1479   .1461
                                                    .1439   .1412   .1381   .1334   .1293   .1238   .1180   .1125   .1066   .1019
                                                    .0981   .0956   .0943   .0936   .0937   .0947   .0961   .0976   .0995   .1017
                                                    .1034   .1047   .1064   .1082   .1095   .1108   .1119   .1127   .1133   .1141
```

```
                                .1147   .1153   .1163   .1172   .1180   .1189   .1193   .1198   .1190   .1197
                                .1176   .1147   .1136   .1100   .1077   .1022   .0992   .0945   .0947   .0904
                                .0911   .0928   .0876   .0900   .0943
Star card                       ********************************************************************************
next                            A4UAU4 IN TMTB, 1 CM CELL    0 DEG
    spectrum                    A4UAU40    350.000    205.000     10.000      8.000   9.390000E-01   1.020000E+02
        info                    .1086   .1085   .1084   .1085   .1095   .1084   .1085   .1083   .1083   .1084
                                .1083   .1083   .1083   .1083   .1082   .1082   .1082   .1081   .1081   .1080
                                .1081   .1078   .1078   .1081   .1078   .1080   .1081   .1078   .1078   .1078
                                .1078   .1078   .1080   .1078   .1078   .1078   .1078   .1078   .1078   .1078
                                .1078   .1078   .1077   .1078   .1078   .1077   .1078   .1080   .1080   .1080
                                .1078   .1080   .1080   .1082   .1084   .1086   .1089   .1096   .1103   .1111
                                .1125   .1135   .1152   .1171   .1190   .1212   .1226   .1246   .1261   .1273
                                .1284   .1299   .1316   .1333   .1350   .1371   .1390   .1409   .1425   .1438
                                .1454   .1465   .1476   .1482   .1490   .1493   .1490   .1486   .1469   .1460
                                .1437   .1407   .1377   .1340   .1288   .1232   .1180   .1122   .1063   .1021
                                .0981   .0958   .0939   .0937   .0936   .0948   .0959   .0978   .0997   .1013
                                .1035   .1050   .1066   .1080   .1097   .1100   .1114   .1121   .1131   .1136
                                .1144   .1155   .1161   .1168   .1172   .1187   .1190   .1194   .1194   .1189
                                .1174   .1152   .1127   .1108   .1070   .1030   .1000   .0965   .0931   .0905
                                .0887   .0878   .0918   .0923   .0918
4HSTOP in place of star card terminates execution
                                STOP

                                1234567891111111111222222222233333333344444444445555555555666666666677777777778
                                          01234567890123456789012345678901234567890123456789012345678901234567890

New baselines for spectra to follow can be inserted
at any time by replacing star card with a B card.
Plot card: YES=make printer plots; 310 = max wavelength plotted.
6 = CDMAX; .2 = CDMAX/50; 0 = ISY, if ISY = 1, CDMAX and
scale calculated by program; 1.2 = max - min wavelength/100.
```

```
                                                  12345678911111111112222222222333333333344444444445555555555666666666677777777778
"DW" input to GETNSM                                      01234567890123456789012345678901234567890123456789012345678901234567890

Baseline to be read                               B
                                                  YES       310.       6.        .2          0        1.2
                                                   BASELINE FOR A6U6 IN DW, 1 CM CELL      4/19/72
OD and E are dummies on DW input                  41972B08   350.000   205.000    10.000     5.000   1.000000E+00   1.000000E+00
                                                     .1039   .1039   .1039   .1039   .1039   .1038   .1033   .1038   .1039   .1038
                                                     .1037   .1038   .1037   .1037   .1037   .1036   .1036   .1036   .1035   .1035
                                                     .1034   .1035   .1034   .1033   .1033   .1034   .1033   .1033   .1034   .1031
                                                     .1031   .1033   .1031   .1031   .1031   .1033   .1031   .1033   .1031   .1031
                                                     .1031   .1031   .1031   .1031   .1031   .1031   .1031   .1031   .1031   .1033
                                                     .1031   .1031   .1031   .1033   .1031   .1031   .1033   .1031   .1031   .1031
                                                     .1031   .1031   .1031   .1034   .1033   .1031   .1031   .1031   .1033   .1033
                                                     .1034   .1031   .1031   .1033   .1033   .1033   .1034   .1034   .1034   .1031
                                                     .1033   .1034   .1033   .1031   .1031   .1033   .1031   .1031   .1031   .1031
                                                     .1031   .1031   .1031   .1031   .1031   .1031   .1033   .1031   .1031   .1031
                                                     .1031   .1031   .1031   .1031   .1031   .1031   .1034   .1031   .1033   .1035
                                                     .1036   .1037   .1038   .1039   .1039   .1042   .1045   .1047   .1050   .1052
                                                     .1052   .1053   .1053   .1058   .1058   .1063   .1063   .1064   .1061   .1064
                                                     .1064   .1064   .1064   .1066   .1064   .1066   .1066   .1064   .1066   .1064
                                                     .1064   .1064   .1074   .1068   .1067
"DW" replaces star card                           DW
                                                   A6U6  IN DW AT 40 DEG 1 CM CELL
OD and E used for calculating CD                  A6U6419    350.000   205.000    10.000     5.000   9.740000E-01   1.100000E+02
                                                     .1038   .1039   .1039   .1038   .1038   .1038   .1038   .1038   .1038   .1039
                                                     .1037   .1036   .1036   .1037   .1038   .1035   .1034   .1034   .1034   .1034
                                                     .1033   .1034   .1034   .1031   .1031   .1031   .1031   .1031   .1031   .1031
                                                     .1031   .1031   .1031   .1031   .1031   .1031   .1031   .1030   .1030   .1031
                                                     .1030   .1030   .1028   .1031   .1031   .1031   .1031   .1031   .1031   .1031
                                                     .1030   .1030   .1031   .1031   .1031   .1031   .1033   .1033   .1036   .1038
                                                     .1041   .1046   .1053   .1060   .1067   .1078   .1093   .1103   .1115   .1130
                                                     .1143   .1151   .1161   .1172   .1180   .1190   .1201   .1210   .1216   .1219
                                                     .1226   .1227   .1229   .1223   .1219   .1214   .1205   .1191   .1176   .1149
                                                     .1125   .1093   .1068   .1033   .1003   .0972   .0945   .0923   .0905   .0897
                                                     .0882   .0880   .0883   .0882   .0891   .0896   .0909   .0920   .0931   .0945
                                                     .0956   .0967   .0980   .0987   .0998   .1010   .1017   .1027   .1035   .1042
                                                     .1047   .1050   .1060   .1070   .1077   .1086   .1097   .1113   .1122   .1125
                                                     .1133   .1122   .1117   .1107   .1086   .1053   .1022   .0992   .0958   .0934
                                                     .0916   .0879   .0884   .0881   .0873
                                                  STOP

                                                  12345678911111111112222222222333333333344444444445555555555666666666677777777778
More DW input in following example.                       01234567890123456789012345678901234567890123456789012345678901234567890
```

"B2" input to GETNSM

"B2" indicates two baselines in a row to be averaged.

OD and E here would be used if normal input were to follow.

second baseline

"DW" says disregard old OD and E

spectrum

```
1234567891111111111222222222233333333334444444444555555555566666666667777777778
012345678901234567890123456789012345678901234567890123456789012345678901234567890
B2
YES      310.      6.      .2        0      1.2
 BASELINE FOR A5U5 AND A4U4 IN DW, 1 CM CELL, .04 SCALE   4/4/72
BL447204   350.000   205.000    10.000     5.000  1.000000E+00   0.
  .0415   .0414   .0413   .0413   .0412   .0411   .0412   .0412   .0411   .0410
  .0410   .0410   .0410   .0408   .0410   .0410   .0409   .0410   .0408   .0407
  .0407   .0406   .0407   .0405   .0406   .0406   .0405   .0406   .0406   .0405
  .0403   .0404   .0404   .0403   .0404   .0404   .0404   .0403   .0403   .0403
  .0401   .0403   .0403   .0403   .0401   .0403   .0402   .0402   .0402   .0402
  .0401   .0401   .0404   .0404   .0403   .0400   .0402   .0400   .0401   .0401
  .0400   .0400   .0399   .0402   .0399   .0398   .0399   .0397   .0397   .0397
  .0398   .0396   .0395   .0393   .0395   .0394   .0393   .0394   .0394   .0394
  .0392   .0394   .0392   .0392   .0391   .0392   .0390   .0390   .0390   .0388
  .0388   .0389   .0389   .0382   .0385   .0385   .0383   .0384   .0381   .0375
  .0378   .0379   .0378   .0375   .0376   .0378   .0378   .0375   .0377   .0378
  .0378   .0376   .0381   .0379   .0383   .0383   .0398   .0389   .0396   .0394
  .0398   .0397   .0397   .0401   .0404   .0398   .0402   .0401   .0406   .0406
  .0403   .0407   .0399   .0400   .0404   .0401   .0402   .0401   .0405   .0402
  .0411   .0407   .0410   .0403   .0415
**********************************************************************************
 SECOND BASELINE FOR A5U5 AND A4U4 IN DW, 1CM CELL, .04 SCALE  4/4/72
447204     350.000   205.000    10.000     5.000  1.000000E+00   1.000000E+00
  .0415   .0414   .0413   .0412   .0413   .0413   .0411   .0410   .0411   .0411
  .0410   .0411   .0410   .0411   .0410   .0410   .0409   .0410   .0408   .0407
  .0406   .0406   .0407   .0405   .0405   .0405   .0403   .0406   .0405   .0405
  .0403   .0404   .0403   .0403   .0403   .0403   .0402   .0402   .0402   .0401
  .0402   .0400   .0402   .0403   .0401   .0401   .0403   .0401   .0401   .0399
  .0402   .0400   .0400   .0399   .0399   .0399   .0398   .0399   .0399   .0399
  .0399   .0399   .0399   .0399   .0395   .0397   .0396   .0393   .0397   .0396
  .0396   .0395   .0397   .0395   .0394   .0391   .0394   .0395   .0393   .0393
  .0393   .0392   .0391   .0392   .0392   .0388   .0389   .0390   .0388   .0388
  .0388   .0388   .0388   .0385   .0385   .0383   .0385   .0384   .0382   .0382
  .0381   .0375   .0375   .0377   .0379   .0378   .0378   .0381   .0378   .0384
  .0384   .0386   .0388   .0384   .0388   .0389   .0386   .0395   .0397   .0397
  .0398   .0405   .0406   .0407   .0410   .0406   .0406   .0408   .0409   .0409
  .0412   .0409   .0410   .0410   .0407   .0407   .0410   .0411   .0421   .0415
  .0418   .0416   .0415   .0419   .0417
DW
 A5U5 IN DW AT 40 DEG  (.04 FULL SCALE), 1 CM CELL
A5U54004   350.000   205.000    10.000     5.000  9.539999E-01  1.110000E+02
  .0414   .0414   .0414   .0411   .0412   .0413   .0413   .0411   .0411   .0413
  .0411   .0413   .0410   .0411   .0410   .0410   .0410   .0409   .0409   .0408
  .0407   .0407   .0407   .0409   .0409   .0407   .0406   .0407   .0407   .0406
  .0407   .0406   .0405   .0405   .0404   .0404   .0404   .0403   .0403   .0403
  .0401   .0403   .0403   .0404   .0402   .0401   .0401   .0403   .0403   .0403
  .0402   .0405   .0403   .0401   .0401   .0403   .0404   .0404   .0406   .0409
```

```
 .0411  .0415  .0421  .0427  .0432  .0443  .0450  .0463  .0476  .0483
 .0490  .0503  .0511  .0519  .0528  .0536  .0547  .0554  .0558  .0554
 .0559  .0562  .0563  .0560  .0556  .0548  .0534  .0517  .0504  .0497
 .0464  .0433  .0410  .0377  .0352  .0327  .0307  .0281  .0272  .0256
 .0251  .0244  .0247  .0252  .0258  .0268  .0276  .0282  .0292  .0303
 .0310  .0329  .0330  .0341  .0350  .0359  .0365  .0374  .0378  .0384
 .0396  .0400  .0406  .0409  .0425  .0430  .0435  .0446  .0455  .0457
 .0464  .0457  .0453  .0428  .0414  .0402  .0368  .0341  .0317  .0285
 .0277  .0253  .0237  .0276  .0251
```

star card says continue what you were doing, so next input will also be treated as "DW" input.
OD and E here used for this spectrum.

```
**********************************************************************
 A4U4 AT 40 DEG IN DW (.04 FULL SCALE)
A4U44ODW   350.000   205.000    10.000     5.000   8.900001E-01  1.114000E+02
 .0417  .0417  .0415  .0417  .0415  .0415  .0414  .0415  .0415  .0414
 .0413  .0413  .0414  .0414  .0414  .0413  .0413  .0412  .0411  .0410
 .0411  .0411  .0412  .0410  .0408  .0408  .0409  .0409  .0409  .0408
 .0410  .0406  .0407  .0407  .0407  .0406  .0409  .0406  .0406  .0406
 .0406  .0406  .0407  .0407  .0407  .0407  .0408  .0405  .0406  .0406
 .0406  .0406  .0406  .0406  .0405  .0406  .0407  .0407  .0407  .0408
 .0412  .0415  .0417  .0421  .0429  .0435  .0442  .0452  .0459  .0467
 .0472  .0480  .0486  .0497  .0501  .0509  .0515  .0516  .0520  .0528
 .0527  .0527  .0526  .0528  .0519  .0514  .0504  .0491  .0482  .0453
 .0443  .0418  .0401  .0377  .0356  .0334  .0324  .0303  .0293  .0281
 .0273  .0276  .0268  .0271  .0278  .0287  .0291  .0301  .0309  .0315
 .0323  .0330  .0338  .0343  .0351  .0362  .0369  .0374  .0383  .0384
 .0391  .0395  .0403  .0411  .0415  .0422  .0431  .0433  .0440  .0446
 .0451  .0445  .0432  .0428  .0420  .0402  .0385  .0352  .0352  .0317
 .0311  .0299  .0298  .0303  .0284
```

&HSTOP terminates execution.

```
$STOP
1234567891111111111222222222233333333334444444444555555555566666666677777777778
          012345678901234567890123456789012345678901234567890123456789012345678890
```

"R-B" input to GETNSM

```
1234567891111111111222222222233333333334444444444555555555566666666677777777778
          012345678901234567890123456789012345678901234567890123456789012345678901234567890
```

Baseline follows

```
B
YES       310.       6.       .2          0      1.2
 BASELINE FOR A2CGU2 RUNS IN .126 MM CELL 8/9/72
8 9 728L    350.000    195.000     10.000     8.000      .904          80.0
   .1041    .1041    .1041    .1039    .1039    .1039    .1041    .1042    .1042    .1042
   .1042    .1042    .1042    .1042    .1042    .1044    .1044    .1042    .1042    .1041
   .1044    .1044    .1044    .1044    .1045    .1045    .1045    .1045    .1045    .1046
   .1046    .1046    .1047    .1046    .1047    .1048    .1047    .1048    .1047    .1048
   .1048    .1048    .1048    .1048    .1047    .1048    .1047    .1048    .1048    .1048
   .1049    .1049    .1048    .1048    .1048    .1047    .1049    .1048    .1049    .1049
   .1048    .1049    .1049    .1050    .1050    .1050    .1050    .1052    .1050    .1050
   .1052    .1053    .1053    .1053    .1053    .1053    .1056    .1056    .1055    .1056
   .1057    .1057    .1058    .1058    .1059    .1059    .1060    .1061    .1059    .1061
   .1063    .1063    .1063    .1066    .1067    .1064    .1067    .1064    .1067    .1068
   .1070    .1073    .1072    .1073    .1075    .1075    .1077    .1077    .1081    .1082
   .1085    .1085    .1088    .1089    .1095    .1097    .1104    .1103    .111C    .1108
   .1111    .1111    .1121    .1160    .1127    .1125    .1127    .1131    .1133    .1133
   .1133    .1133    .1136    .1136    .1141    .1140    .1136    .1133    .1140    .1139
   .1149    .1138    .1152    .1140    .1143    .1136    .1142    .1136    .1147    .1136
   .1115    .1110    .1102    .1095    .1084
```

"R-B" replaces star card, says following spectra punched as spectrum - baseline.
OD and E must be the same as when the tapes were punched.

```
R-B
 A2CGU2 AT -3.5 DEG IN .126 MM CELL
A2CGU2-4    350.000    195.000     10.000     8.000   1.080000E+00   9.630000E+01
  -.0446   -.0535   -.0446   -.0446   -.0446   -.0446   -.0535   -.0714   -.0624   -.0803
  -.0624   -.0714   -.0714   -.0714   -.0803   -.0981   -.0981   -.0803   -.0892   -.0714
  -.0981   -.0981   -.0981   -.0981   -.1249   -.1160   -.1160   -.1160   -.1249   -.1427
  -.1338   -.1427   -.1694   -.1694   -.1873   -.2141   -.2052   -.2230   -.2319   -.2588
  -.2766   -.3032   -.3210   -.3210   -.3745   -.4014   -.4370   -.4639   -.5176   -.5620
  -.6064   -.6421   -.6602   -.6602   -.6421   -.6064   -.5352   -.4104   -.2230   -.0624
   .1960    .4458    .7485    .9985   1.3018   1.5332   1.8008   1.9795   2.1660   2.2988
  2.4688   2.6289   2.8340   3.0313   3.2617   3.4238   3.6016   3.7969   3.9844   4.0858
  4.1719   4.2227   4.1875   4.1250   3.9766   3.8066   3.5117   3.3242   3.0215   2.7266
  2.3887   2.0586   1.6759   1.3281    .9536    .7310    .4290    .2764    .0624   -.0357
  -.1338   -.2230   -.2588   -.2854   -.3479   -.3745   -.3479   -.4014   -.3567   -.4104
  -.4104   -.3479   -.3032   -.3032   -.2676   -.1783   -.1694   -.0357   0.       .1337
   .1960    .3833    .4368    .5703    .6865    .7485    .8823    .8467    .8110    .6152
   .4280    .2673   -.1963   -.4014   -.8921  -1.2754  -1.6680  -2.0957  -2.4434  -2.6055
 -2.9082  -3.0332  -3.4160  -3.0508  -3.0781  -3.1484  -2.9434  -2.9160  -2.4082  -2.1504
 -1.7305  -1.0166  -1.0264   -.3301    .1426
```

star card says continue what you were doing, so next input will also be treated as "R-B" input.
OD and E must be the same as when the tapes were punched.

```
**********************************************************************************
 0 DEG
A2CGU20     350.000    195.000     10.000     8.000   1.080000E+00   9.630000E+01
  -.0268   -.0268   -.0357   -.0178   -.0178   -.0268   -.0268   -.0357   -.0357   -.0446
  -.0446   -.0535   -.0535   -.0357   -.0624   -.0714   -.0624   -.0535   -.0624   -.0357
  -.0714   -.0714   -.0803   -.0803   -.0892   -.0803   -.0892   -.0892   -.1071   -.1071
  -.1071   -.1160   -.1249   -.1249   -.1427   -.1605   -.1605   -.1783   -.1873   -.2141
  -.2319   -.2598   -.2676   -.2766   -.3210   -.3389   -.3745   -.4192   -.4548   -.4995
  -.5352   -.5708   -.5977   -.5977   -.5889   -.5352   -.4905   -.3479   -.1783   0.
   .2496    .4636    .7842   1.0156   1.3369   1.5596   1.8184   1.9795   2.1660   2.3262
  2.4863   2.6465   2.8516   3.0391   3.2539   3.4590   3.6465   3.8066   3.9766   4.1016
  4.2070   4.2695   4.2148   4.1641   4.0000   3.8516   3.6191   3.3613   3.0840   2.7363
  2.4160   2.1211   1.6934   1.4082    .9985    .7129    .4636    .3208    .1159   0.
  -.1338   -.2052   -.2408   -.2854   -.3301   -.3210   -.3389   -.3389   -.3567   -.3389
  -.3657   -.3210   -.2854   -.2588   -.1962   -.1605   -.1160   -.0268    .0535    .1604
   .2942    .4189    .4993    .6152    .7041    .8379    .9897    .9064    .8467    .7397
   .5815    .4011    .0446   -.2141   -.8652  -1.2314  -1.5518  -1.7930  -2.2832  -2.5156
 -2.7285  -2.9160  -3.3006  -2.9707  -3.2637  -2.9883  -2.8008  -2.8633  -2.1504  -2.4180
 -1.3740  -.8115  -1.3379   -.0268    .2764
STOP
```

&HSTOP terminates execution.

```
1234567891111111111222222222233333333334444444444555555555566666666677777777778
          012345678901234567890123456789012345678901234567890123456789012345678901234567890
```

E. Program CDCONT

This program performs the procedures described in Chapter IV.A.1. However, some of the variable names are different:

Chapter IV		CDCONT
n	=	CHNLTH
t	=	NPTS
v	=	NSPC
ℓ	=	NN
$\underset{\sim}{Y}$	=	CMON
$\underset{\sim}{Q}$	=	X
$(\underset{\sim}{Q}\cdot\underset{\sim}{C}+\underset{\sim}{Y}\cdot\underset{\sim}{E})$	=	SP.

The order of first neighbor contributions on input is the same as described in Chapter I.B.4.b.1; the monomers are in the order A,C,G,U. On input the monomer coefficients, spectra, and $\underset{\sim}{Q}$ matrix elements must all read down the page in the same order. In the example this means A_4U_4 first, A_6U_6 second, A_4UAU_4 third, ..., $A_4CG + CGU_4$ last. The listing and sample input follow.

Unlisted routines required:

PRNT	(see SMOOSP listing)
PRNPLT	(see SMOOSP listing)
PLSCAL	(see SMOOSP listing)

0 0 0 0 3 9 0 0 9 3 6


```
      PROGRAM CDCONT(INPUT,OUTPUT,PUNCH)
C
C  THE MATRIX, Q, DESCRIBED IN THESIS IS EQUIVALENT TO MATRIX X HERE
C
      COMMON/VAR/ S(30,200), SP(30,200), X(30,16), H(16,16), HI(16,16) ,
     1 C(30,20 ), FNAME(30,4) , CON(7), E(4,200), CMON(30,4), CC(16),
     1 W(2 ), XT(16,3 ),TMP(16,30),CHNLTH(30)
      COMMON/VAR2/NPTS,NN,NSPC
      DIMENSION  SCR(2000), B(16,16), CNAME(16,4)
      DO 1  I=1,200
      DO 1  J=1,30
      DO 1  K=1,16
      X(J,K)=0.
   1  S(I,J)=0.
      NPTS=100
      READ 904,((E(I,J),J=1,130),I=1,4)
      PRINT 950
      PRINT904,((E(I,J),J=1,130),I=1,4)
      PRINT 950
      READ 900,NSPC,NN
      PRINT 900,NSPC,NN
      DO 17 I=1,NSPC
      READ 734, (CMON(I,J),J=1,4),CHNLTH(I)
  17  PRINT734, (CMON(I,J),J=1,4),CHNLTH(I)
      PRINT 950
      DO 20 I=1,NSPC
      READ 902,(FNAME(I,J),J=1,4)
      PRINT902,(FNAME(I,J),J=1,4)
      READ 903,(CON(J),J=1,7)
      PRINT903,(CON(J),J=1,7)
      XINC=CON(4)/10.
      MPTS=(CON(2)-CON(3))/XINC
      PRINT 90 ,MPTS,NPTS
      READ 904,(S(I,J),J=1,MPTS)
      PRINT904,(S(I,J),J=1,MPTS)
      READ 935,STAR
      PRINT905,STAR
C
C  MAKE LOW WAVELENGTH CUTOFF = THAT OF SPECTRUM WITH FEWEST PTS,
C  ALL MUST START AT 350 NM.
C
      IF(MPTS.LT.NPTS)NPTS=MPTS
   20 CONTINUE
      DO 22 J=1,NSPC
   22 READ 906, (X(J,K),K=1,NN)
      PRINT 910
      PRINT 911
      DO 23 J=1,NSPC
      IF(NN .EQ. 16)GO TO 27
      PRINT 812, (X(J,K),K=1,NN), (FNAME(J,K),K=1,4)
      GO TO 23
   27 PRINT 912, (X(J,K),K=1,NN), (FNAME(J,K),K=1,4)
   23 CONTINUE
      PRINT 914
      MNPTS = NPTS - 30
      DO 36 I=1,4
      PRINT 950
   36 PRINT 904, (E(I,J),J=1,MNPTS)
```

```
C
C THE MONOMER SPECTRA START AT 320 NM
C
      PRINT 915
      DO 38 I=1,NSPC
   38 PRINT 815, (CMON(I,J),J=1,4),CHNLTH(I),(FNAME(I,J),J=1,4)
      DO 80 I=1,NN
      DO 80 J=1,NSPC
   80 XT(I,J) = X(J,I)
C
C  COMPUTE XTRANSPOSE*X (ABBREV. XT*X), X IS THE COEFFICIENT MATRIX.
C
      CALL MATOP(XT,X,H,NN,NSPC,NN,16,30,16,1)
      PRINT 910
      PRINT 917
      DO 81  J=1,NN
   81 PRINT 804, (H(J,K),K=1,NN)
C
C  INVERT (XT*X)   (=HI)
C
      DO 6 K=1,NN
      DO 5 J=1,NN
    5 B(K,J) = 0.
    6 B(K,K) = 1.
      CALL LINIT(H,B,HI,NN,NN,DET,IEX,CNR,SINGUL,16,SCR)
      PRINT 910
      PRINT 930, DET,IEX,CNR
      IF(SINGUL.EQ. .FALSE.) 8,7
    7 PRINT 931
      GO TO 9
    8 PRINT 932
    9 CONTINUE
      PRINT 918
      DO 91 J=1,NN
   91 PRINT 804, (HI(J,L),L=1,NN)
C
C  CHECK INVERSE
C
      CALL MATOP(H,HI,C,NN,NN,NN,16,16,30,1)
      PRINT 919
      DO 101 J=1,NN
  101 PRINT 804, (C(J,K),K=1,NN)
C
C  REDUCE RANGE OF INPUT DATA SO START IS 320 INSTEAD OF 350 NM,
C  AND MULT BY CHAINLENGTH FACTOR
C
      NPTS = MNPTS
      DO 30 I=1,NSPC
      IF(NN.EQ.10) CHNLTH(I) = 2.*CHNLTH(I)
      DO 30 J=1,NPTS
      K=J+30
      SP(I,J) = S(I,K)
   30 S(I,J)=SP(I,J)*CHNLTH(I)
      W(1)=CON(2)-30.
      DO 110 I=2,NPTS
  110 W(I)=W(I-1)-XINC
      PRINT 910
      PRINT 913,NPTS
```

```
      DO 320 I=1,NSPC
      PRINT 910
      PRINT 902, (FNAME(I,J), J=1,4)
      PRINT 950
      DO 315 J=1,NPTS
  315 C(J) = SP(I,J)
      PRINT 960
      CALL PRNT(W,C,NPTS)
      PRINT 910
      PRINT 902,(FNAME(I,J) , J=1,4)
      PRINT 950
      CALL PRNPLT(W,C,320.,1.3, 6.,.2 ,0,2,NPTS)
  320 CONTINUE
C
C  GET END CORRECTION FROM MONOMER SPECTRA AND MONOMER COEFFICIENTS.
C
      CALL MATOP(CMON,E,SP,NSPC,4,NPTS,30,4,30,1)
      DO 50 I=1,NSPC
      DO 50 J=1,NPTS
   50 SP(I,J) = S(I,J) - SP(I,J)
C
C  THIS IS THE S PRIMED MATRIX
C
      PRINT 910
      PRINT 916
      DO 51 L=1,NSPC
      PRINT 950
   51 PRINT 704, (SP(L,J), J=1,NPTS)
C
C  CALCULATE THE FIRST NEIGHBOR CD CONTRIBUTIONS  (=(XT*X)INV * XT * SP)
C
      CALL MATOP(HI,XT,TMP,NN,NN,NSPC,16,16,16,1)
      CALL MATOP(TMP,SP,C,NN,NSPC,NPTS,16,30,30,1)
      PRINT 910
      PRINT 945
      DO 57 I=1,NSPC
   57 PRINT 902, (FNAME(I,J), J=1,4)
      DO 58 I=1,NN
      PRINT 950
   58 PRINT 904, (TMP(I,J), J=1,NSPC)
      PRINT 910
      PRINT 920
      CC(1)=2HAA
      CC(2)=2HAC
      CC(3)=2HAG
      CC(4)=2HAU
      CC(5)=2HCA
      CC(6)=2HCC
      CC(7)=2HCG
      IF(NN.EQ.16) GO TO 150
      CC(8)=2HGA
      CC(9)=2HGC
      CC(10)=2HUA
      SOD=2HDS
      GO TO 170
  150 CC(8)=2HCU
      CC(9)=2HGA
      CC(10)=2HGC
```

```
      CC(11)=2HGG
      CC(12)=2HGU
      CC(13)=2HUA
      CC(14)=2HUC
      CC(15)=2HUG
      CC(16)=2HUU
      SOD = 2HSS
  170 CONTINUE
      CON(2) = W(1)
      CON(3) = W(1) - NPTS*XINC
      DO 180 J=1,NN
      DO 175 I=1,NPTS
  175 S(I) = C(J,I)
      PRINT 910
      PRINT 921, SOD, CC(J)
      PRINT 960
      CALL PRNT(W,S,NPTS)
      PRINT 910
      PRINT 921,SOD,CC(J)
      CALL PRNPLT(W,S,320.,1.3,60., 2.    ,0.0,NPTS)
      CON(1) = 8H
       CNAME(J,1) = CC(J)
       CNAME(J,2) = SOD
       CNAME(J,3) = 10HCONTRIBUTI
       CNAME(J,4) = 10HON
      PUNCH 902, (CNAME(J,K), K=1,4)
      CON(1) = CC(J)
      PUNCH 903, (CON(K), K=1,7)
      PUNCH 904, (C(J,K), K=1,NPTS)
      PUNCH 980
  180 CONTINUE
C
C  RECONSTITUTE THE INPUT VECTORS FROM X AND C
C
      IF(NN .NE. 10) GO TO 182
      DO 181 I=1,NSPC
  181 CHNLTH(I) = .5 * CHNLTH(I)
  182 CONTINUE
      INDX = 6HDSOLIG
      CALL CTOS(INDX,NSPC)
C
C REMOVE NEXT CALL IF YOU DO NOT WANT TO CALCULATE MORE SPECTRA FROM THE C*S.
C
      CALL FREADER
  700 FORMAT(18F7.4)
  900 FORMAT(515)
  901 FORMAT(A3,7X,3F10.4,I5,F15.4)
  902 FORMAT(4A10)
  903 FORMAT(A8,2X,4(F8.3,2X),2(E13.6,2X))
  899 FORMAT(1X,2(A2,1X),3X,2F10.3)
  904 FORMAT(10F8.4)
  804 FORMAT(16F8.3)
  704 FORMAT(10F10.4)
  905 FORMAT(A5)
  906 FORMAT(16F5.1)
  910 FORMAT(1H1)
  911 FORMAT(* THIS IS THE X MATRIX, THE ROWS ARE THE NUMBER OF EACH TYP
     1E OF INTERACTION IN THE INPUT SPECTRA*/
```

0 0 0 0 3 9 0 0 2 6 8


```
     180H  AA   AC   AG   AU   CA   CC   CG   CU   GA   GC   GG   GU   U
     1A   UC   UG   UU ,* FOR SINGLE STRANDS*/
     180H  AA   AC   AG   AU   CA   CC   CG   GA   GC   UA
     1                  ,* FOR DOUBLE STRANDS*/)
  812 FORMAT(10F5.1* FOR *4A10)
  912 FORMAT(16F5.1* FOR *4A10)
  913 FORMAT(* THESE ARE THE INPUT SPECTRA TRUNCATED TO*I5* WAVELENGTHS*
     1/)
  914 FORMAT(///* THESE ARE THE MONOMER SPECTRA IN THE ORDER A,C,G,U*/)
  915 FORMAT(///* THESE ARE THE MONOMER COEFFICIENTS FOR THE END BASES*/
     1*  A     C     G     U     CHAINLENGTH*/)
  815 FORMAT(4F5.1,F10.1,5X,* FOR *,4A10/)
  916 FORMAT(///* THIS IS THE S PRIMED MATRIX IT CONTAINS THE INPUT SPEC
     1TRA*/* WITH CHAINLENGTH WEIGHTING AND MONOMER CORRECTIONS FOR ENDS
     2*/)
  917 FORMAT(///* THIS IS XTRANSPOSE TIMES X(=H),IS IT SQUARE SYMMETRIC*
     1/)
  918 FORMAT(///* THIS IS H INVERSE*/)
  919 FORMAT(///* THIS IS H TIMES H INVERSE, DOES IT = UNIT MATRIX*/)
  920 FORMAT(* ALL CHECKS COMPLETE, FIRST NEIGHBOR CONTRIBUTIONS FOLLOW*
     1)
  921 FORMAT(1X,2(A2,1X),* CONTRIBUTION*)
  922 FORMAT(80(1H*))
  930 FORMAT(* DET=*,E16.7,* IEX=*,I5,* CNR=*,E16.7)
  931 FORMAT(//* THE MATRIX IS SINGULAR*)
  932 FORMAT(//* THE MATRIX IS NOT SINGULAR*)
  950 FORMAT(/)
  945 FORMAT(34H THIS IS THE MATRIX (XT*X)INV * XT,/ * THE SPACES SEPARA
     1TE ROWS.  THE ELEMENTS OF THE FIRST ROW ARE THE COEFFICIENTS USED
     1TO CONSTRUCT THE AA CONTRIBUTION.*/* THE CONTRIBUTIONS ARE THE SUM
     1 OF ROW ELEMENTS TIMES THE APPROPRIATE S PRIMED SPECTRUM.*/
     1 * THE ORDER OF SPECTRA IN S PRIME IS,*/)
  960 FORMAT(/,12X,3(* WAVELENGTH*, 6X,* PROPERTY*,10X),/)
  980 FORMAT(80(1H*))
      STOP
      END
```

```
      SUBROUTINE CTOS(INDX,NCALC)
C
C  SUBROUTINE TO GO FROM C MATRIX TO SPECTRA.
C INDX,NCALC, FNAME, F, CMON, AND CHNLTH MUST BE SPECIFIED IN MAIN
C  PROGRAM OR READ IN VIA SUBROUTINE FREADER.
C
      COMMON/VAR/ S(30,200), SP(30,200), F(30,16), H(16,16), HI(16,16) ,
     1 C(30,200), FNAME(30,4) , CON(7), E(4,200), CMON(30,4), CC(16),
     1 W(200), XT(16,30),TMP(16,30),CHNLTH(30)
      COMMON/VAR2/NPTS,NN,NSPC
      DO 40 I=1,NCALC
      IF(INDX.NE.6HDSPOLY) GO TO 10
      DO  9 J=1,NN
    9 F(I,J) = .5*F(I,J)
   10 CONTINUE
      IF(INDX.NE.6HDSOLIG) GO TO 20
      DO 18 K=1,4
   18 CMON(I,K) = .5*CMON(I,K)/CHNLTH(I)
      DO 19 J=1,NN
   19 F(I,J) = .5*F(I,J)/CHNLTH(I)
   20 CONTINUE
      IF(INDX.NE.6HSSOLIG) GO TO 40
      DO 38  K=1,4
   38 CMON(I,K) = CMON(I,K)/CHNLTH(I)
      DO 39 J=1,NN
   39 F(I,J)=F(I,J)/CHNLTH(I)
   40 CONTINUE
      PRINT 960
      CALL MATOP(F,C,SP,NCALC,NN,NPTS,30,30,30,1)
      CALL MATOP(CMON,E,S,NCALC,4,NPTS,30,4,30,1)
      DO 45 I=1,NCALC
      DO 45 J=1,NPTS
   45 S(I,J) = S(I,J) + SP(I,J)
      DO 70 I=1,NCALC
      PRINT 910
      PRINT 940
      PRINT 902, (FNAME(I,J), J=1,4)
      PRINT 960
      DO 50 J=1,NPTS
   50 SP(J) = S(I,J)
      CALL PRNT(W,SP,NPTS)
      PRINT 910
      PRINT 940
      PRINT 902, (FNAME(I,J), J=1,4)
      CALL PRNPLT(W,SP,320.,1.3, 6.,.2 ,0,0,NPTS)
      CON(1) = 8HCONSTRCT
      PUNCH 902, (FNAME(I,J), J=1,4)
      PUNCH 903, (CON(J), J=1,7)
      PUNCH 904, ( S(I,J), J=1,NPTS)
      PUNCH 980
   70 CONTINUE
      RETURN
  902 FORMAT(4A10)
  903 FORMAT(A8,2X,4(F8.3,2X),2(E13.6,2X))
  904 FORMAT(10F8.4)
  910 FORMAT(1H1)
  940 FORMAT(* THIS SPECTRUM IS CONSTRUCTED FROM X,C,CMON, AND E FOR*/)
  960 FORMAT(/,12X,3(* WAVELENGTH*, 6X,* PROPERTY*,10X),/)
  980 FORMAT(80(1H*))
      END
```

0 0 0 0 3 9 0 0 9 6 9


```
      SUBROUTINE FREADER
C
C  NSETS = NO SETS OF SPECTRA TO BE CALCULATED
C  INDX= DSPOLY, DSOLIGO, SSPOLY, OR SSOLIGO .  THESE HOLLERITH CONSTANTS
C  MUST START IN COLUMN 1, THEY INDICATE THE KIND OF DATA IN THE INPUT.
C  NCALC IS THE NUMBER OF SPECTRA IN THE SET
C FNAME IS A 40 CHARACTER OR LESS NAME
C F IS THE NUMBER OF INTERACTIONS IN THE ORDER AA, AC,AG, ETC.  16 PER
C  CARD. THEY ARE THE FIRST NEIGHBOR FREQUENCIES FOR POLYMERS.
C CMON ARE THE MONOMER COEFFICIENTS TO ACCOUNT FOR THE ENDS, CHNLTH IS
C THE CHAINLENGTH.  THESE CARDS ARE LEFT BLANK FOR POLYMERS
C
      COMMON/VAR/ S(30,200), SP(30,200), F(30,16), H(16,16), HI(16,16) ,
     1 C(30,200), FNAME(30,4) , CON(7), E(4,200), CMON(30,4), CC(16),
     1 W(200), XT(16,30),TMP(16,30),CHNLTH(30)
      COMMON/VAR2/NPTS,NN,NSPC
      READ 900,NSETS
      DO 100 J=1,NSETS
      READ 800, INDX,NCALC
      DO 40 I=1,NCALC
      READ 902, (FNAME(I,J), J=1,4)
      READ 906, (F(I,J) , J=1,NN)
      READ 704, (CMON(I,J), J=1,4), CHNLTH(I)
   40 CONTINUE
      CALL CTOS(INDX,NCALC)
  100 CONTINUE
  800 FORMAT(A6,4X, I5)
  900 FORMAT(5I5)
  902 FORMAT(4A10)
  906 FORMAT(16F5.1)
  704 FORMAT(10F10.4)
      RETURN
      END

      SUBROUTINE LINIT(A,B,X,N,M,DET,EX,CNR,SINGUL,L,SCR)
      DIMENSION A(L,1),B(L,1),X(L,1),SCR(L,1)
C     MOVE A AND B TO SCRATCH AREA
      N=MIND(N,L)
      IF (N .LT. 1) GO TO 4
      DO 3 I=1,N
      DO 1 J=1,N
      SCR(I,J)=A(I,J)
    1 CONTINUE
      IF (M .LT. 1) GO TO 3
      DO 2 J=1,M
      K=2*N+4+J
      SCR(I,K)=B(I,J)
    2 CONTINUE
    3 CONTINUE
    4 CONTINUE
      CALL LINEAR(SCR(1,1),SCR(1,2*N+5),X,N,M,DET,EX,CNR,SINGUL,L
     X,SCR(1,N+1),SCR(1,2*N+1),SCR(1,2*N+2),SCR(1,2*N+3),SCR(1,2*N+4))
      RETURN
      END
```

```
      SUBROUTINE LINEAR(A,B,X,N,M,DET,EX,CNR,SINGUL,L
     X,LU,PIVOT,Y,RES,MULT)
C PROCEDURE *LINEAR SYSTEM*
C     6600    6600    6600    6600    6600    6600    6600    6600
C     N=ORDER, B=RH SIDES, M=NO OF RH SIDES, X=ANSWERS,
C     (DET*10**EX)=DETERMINANT, CNR=CONDITION, SINGUL=.FALSE. IF OK.
      INTEGER EX,PIVOT
      REAL LU ,MULT
      LOGICAL SINGUL
      DIMENSION A(L,1),B(L,1),X(L,1),LU(L,1)
     X,PIVOT(1),Y(1),RES(1),MULT(1)
C.....LINEAR IS BASED ON ALGORITHM 135, CACM NOVEMBER 1962 PAGE 553,
C     AS CORRECTED CACM JULY 1964 PAGE 421.
C     IT USES CROUT*S METHOD WITH ROW EQUILIBRATION, ROW INTERCHANGES,
C     AND ITERATIVE IMPROVEMENT FOR SOLVING THE MATRIX EQUATION AX=B,
C     WHERE A IS N BY N, X AND B ARE N BY M.  IN CASE M .LE. 0, ONLY THE
C     DETERMINANT OF A IS EVALUATED.  FOR M=1, THE SUBROUTINE SOLVES A
C     SYSTEM OF LINEAR EQUATIONS IN N UNKNOWNS.  FOR M=N AND B=(THE
C     IDENTITY MATRIX), X IS SET TO THE INVERSE OF A.
C     IF A IS NEARLY SINGULAR, *SINGUL* IS SET TO .TRUE.
C
      DATA EPS / O 1640777777777777777 /
      SINGUL=.FALSE.
      CNR=1.0
      DET=0.0
      EXP=0
      IF ((N .GE. 1) .AND. (N .LE. L)) GO TO 1000
      SINGUL=.TRUE.
      RETURN
 1000 CONTINUE
C
C     REMOVE APPROPRIATE FACTORS FROM THE ROWS OF A.
C     CALL *EQUILIBRATE*
      CALL EQUILI(A,N,MULT,SINGUL,L)
      IF(SINGUL) RETURN
C     SAVE THE RESULT FOR COMPUTATION OF RESIDUALS DURING ITERATION.
      DO 1002 I=1,N
      DO 1001 J=1,N
      LU(I,J)=A(I,J)
 1001 CONTINUE
 1002 CONTINUE
C     DECOMPOSE THE MATRIX INTO TRIANGULAR FACTORS.
C     CALL *CROUT*
      CALL CROUT(LU,N,PIVOT,DET,SINGUL,L)
      IF (SINGUL) RETURN
C     EVALUATE THE DETERMINANT IN THE FORM (DET*10.0**EX)
      DO 1003 I=1,N
      Y(I)=LU(I,I)*MULT(I)
 1003 CONTINUE
C     *PRODUCT*
      DET=DET*PRODUC(Y,1,N,EX)
C     NOW BEGIN TO PROCESS RIGHT HAND SIDES.
      IF (M .LT. 1) RETURN
      DO 1004 K=1,M
      RK=FLOAT(K)
C     SCALE THE RIGHT HAND SIDES
      DO 1005 I1=1,N
      RES(I1)=B(I1,K)/MULT(I1)
```

```
      B(I1,K)=RES(I1)
 1005 CONTINUE
C     STORE THE FIRST APPROXIMATION AND ITS NORM.
      YNORM=0.)
C     CALL *SOLVE*
      CALL SOLVE(LU,N,RES,PIVOT,Y,L)
      DO 1006 I1=1,N
      YNORM=YNORM+ABS(Y(I1))
      X(I1,K)=Y(I1)
 1006 CONTINUE
C     BEGIN THE ITERATION LOOP.  THE NUMBER OF ITERATIONS IS DETERMINED
C     DURING THE FIRST ITERATION.
      KOUNT=1
 1012 CONTINUE
C     COMPUTE THE RESIDUALS OF THE SOLUTION Y.
C     CALL *RESIDUALS*
      CALL RESIDU(A,N,B,K,X,RES,L)
C     FIND NEXT INCREMENT TO THE SOLUTION
C     CALL *SOLVE*
      CALL SOLVE(LU,N,RES,PIVOT,Y,L)
C     SET UP TERMINATION CONDITIONS
      IF (KOUNT .NE. 1) GO TO 1007
      DYNORM=0.0
      DO 1008 I2=1,N
      DYNORM=DYNORM+ABS(Y(I2))
 1008 CONTINUE
      IF (DYNORM .EQ. 0.0) GO TO 1009
      T=YNORM/DYNORM
C     CNR IS AN APPROXIMATION TO THE SPECTRAL NORM OF A, WHICH IS THE
C     PRODUCT OF THE NORM OF A AND THE NORM OF A-INVERSE.  SEE
C     WILKINSON, JACM JULY 1961 PAGE 281.
      CNR=((RK-1.0)*CNR + 1.0/(EPS*T))/RK
      IF (T .GE. 2.0) GO TO 1010
      SINGUL = .TRUE.
      RETURN
 1010 CONTINUE
      LIMIT=IFIX(ALOG(EPS)/ALOG(1.0/T))
 1007 CONTINUE
C     STORE THE NEW APPROXIMATION
      DO 1011 I2=1,N
      X(I2,K)=X(I2,K)+Y(I2)
 1011 CONTINUE
      KOUNT=KOUNT+1
      IF (KOUNT .LE. LIMIT) GO TO 1012
 1009 CONTINUE
 1004 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE EQUILI(A,N,MULT,SINGUL,L)
C PROCEDURE *EQUILIBRATE*
C     N=ORDER,MULT=MULTIPLIERS
      REAL MULT,MX,MASK
      LOGICAL SINGUL
      DIMENSION A(L,1)
     X,MULT(1)
C.....SCALING THE ROWS OF THE MATRIX (A) TO ROUGHLY THE SAME MAXIMUM
C     MAGNITUDE ALLOWS THE PROCEDURE (CROUT) TO SELECT EFFECTIVE
C     PIVOTAL ELEMENTS FOR GAUSSIAN DECOMPOSITION OF THE MATRIX.  A
C     POWER OF 2 IS USED INSTEAD OF THE ACTUAL LARGEST ELEMENT TO
C     REDUCE ROUNDING ERROR IN THE DIVISION.
C     SEE WILKINSON, JACM JULY 1961 PAGE 284.
      DATA MASK / O 777740000000000000000 /
      DO 2001 I=1,N
      MX=0.0
C     FIND THE LARGEST ELEMENT
      DO 2002 J=1,N
      IF (ABS(A(I,J)) .GT. MX) MX=ABS(A(I,J))
 2002 CONTINUE
      IF (MX .GT. 0.0) GO TO 2003
      SINGUL=.TRUE.
      RETURN
 2003 CONTINUE
C     NOW STORE THE MULTIPLIER AND SCALE THE ROW.
      MULT(I)=AND(MX,MASK)
      IF (MULT(I) .EQ. 1.0) GO TO 2004
      DO 2005 J=1,N
      A(I,J)=A(I,J)/MULT(I)
 2005 CONTINUE
 2004 CONTINUE
 2001 CONTINUE
      RETURN
      END
```

0 0 0 0 3 9 0 0 7 1


```
      SUBROUTINE CROUT(A,N,PIVOT,SG,SINGUL,L)
C PROCEDURE *CROUT*
C     N=ORDER, PIVOT=PIVOTS, SG=INTERCHANGES.
      INTEGER PIVOT,P
      DOUBLE PRECISION D
      LOGICAL SINGUL
      DIMENSION A(L,1)
     X,PIVOT(1)
C.....CROUT*S METHOD WITH ROW INTERCHANGES FOR A=L*U WITH L(K,K)=1.0.
C     (PIVOT) STORES THE PERMUTATION MATRIX.
      SG=1.0
C     K IS THE STAGE OF THE ELIMINATION.
      DO 3001 K=1,N
      T=0.0
      DO 3002 I=K,N
C     COMPUTE L.
      KM1=K-1
      D=DBLE(A(I,K))
      IF (KM1 .LT. 1) GO TO 3003
      DO 3004 P=1,KM1
      D=D-DBLE(A(I,P))*DBLE(A(P,K))
 3004 CONTINUE
 3003 CONTINUE
      A(I,K)=SNGL(D)
      IF (ABS(A(I,K)) .LE. T) GO TO 3005
      T=ABS(A(I,K))
      IMAX=I
 3005 CONTINUE
 3002 CONTINUE
      IF (T .GT. 0.0) GO TO 3105
      SINGUL=.TRUE.
      RETURN
 3105 CONTINUE
C     A(IMAX,K) IS LARGEST ELEMENT IN REMAINDER OF COLUMN K.
C     INTERCHANGE ROWS IF NECESSARY.
      PIVOT(K)=IMAX
      IF (IMAX .EQ. K) GO TO 3006
      SG=-SG
      DO 3007 J=1,N
      T=A(K,J)
      A(K,J)=A(IMAX,J)
      A(IMAX,J)=T
 3007 CONTINUE
 3006 CONTINUE
C     COMPUTE A COLUMN OF MULTIPLIERS.
      QUOT=1.0/A(K,K)
      KP1=K+1
      IF (KP1 .GT. N) GO TO 3109
      DO 3008 I=KP1,N
      A(I,K)=A(I,K)*QUOT
 3008 CONTINUE
C     AND COMPUTE A ROW OF U
      DO 3009 J=KP1,N
      D=DBLE(A(K,J))
      IF (KM1 .LT. 1) GO TO 3010
      DO 3011 P=1,KM1
      D=D-DBLE(A(K,P))*DBLE(A(P,J))
 3011 CONTINUE
 3010 CONTINUE
      A(K,J)=SNGL(D)
 3009 CONTINUE
 3109 CONTINUE
 3001 CONTINUE
      RETURN
      END
```

```
      REAL FUNCTION PRODUC(FACTOR,S,F,EX)
C PROCEDURE #PRODUCT#
C     S=START, F=FINISH, EX=EXPONENT
      INTEGER S,F,EX
      DIMENSION FACTOR(1)
C.....MULTIPLIES FACTOR(S) THROUGH FACTOR(F), CORRECTING TO PREVENT
C     OVERFLOW.  THE ABSOULTE VALUE OF THE RESULT IS BETWEEN 0.1 AND 1.0
C     AND THE EXPONENT APPEARS IN (EX).
      EX=0.0
      P=1.0
      IF (S .GT. F) GO TO 4008
      DO 4001 I=S,F
      P1=FACTOR(I)
      IF (ABS(P1) .GE. 0.1) GO TO 4002
      P1=10.0*P1
      EX=EX-1
 4002 CONTINUE
      P=P*P1
      IF (P .NE. 0.0) GO TO 4003
      EX=0
      GO TO 4008
 4003 CONTINUE
 4004 IF (ABS(P) .GE. 0.1) GO TO 4005
      P=P*10.0
      EX=EX-1
      GO TO 4004
 4005 CONTINUE
 4006 IF (ABS(P) .LT. 1.0) GO TO 4007
      P=P/10.0
      EX=EX+1
      GO TO 4006
 4007 CONTINUE
 4001 CONTINUE
 4008 CONTINUE
      PRODUC=P
      RETURN
      END


      SUBROUTINE RESIDU(A,N,B,K,X,RES,L)
C PROCEDURE #RESIDUALS#
C     N=ORDER, B=RH SIDES, K=COLUMN OF B, X=APPROXIMATE SOLUTION,
C     RES=RESIDUALS
      DIMENSION A(L,1),B(L,1),X(L,1)
     X,RES(1)
      INTEGER P
      DOUBLE PRECISION D
C.....COMPUTES B(.,K)-A*X(.,K)
      DO 5001 I=1,N
      D=DBLE(B(I,K))
      IF (N .LT. 1) GO TO 5003
      DO 5002 P=1,N
      D=D-DBLE(A(I,P))*DBLE(X(P,K))
 5002 CONTINUE
 5003 CONTINUE
      RES(I)=SNGL(D)
 5001 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE SOLVE(A,N,V,PIVOT,Y,L)
C PROCEDURE *SOLVE*
C     N=ORDER, V=RH VECTOR, PIVOT=PERMUTATION VECTOR, Y=ANSWER.
      INTEGER PIVOT,P
      DIMENSION A(L,1)
     X,PIVOT(1),V(1),Y(1)
C.....PROCESSES A RH VECTOR AND THEN BACK-SOLVES FOR Y USING THE L*U
C     DECOMPOSITION PROVIDED BY (CROUT).
      DO 6001 K=1,N
      J=PIVOT(K)
      T=V(J)
      V(J)=V(K)
      KM1=K-1
      IF (KM1 .LT. 1) GO TO 6102
      DO 6002 P=1,KM1
      T=T-A(K,P)*V(P)
 6002 CONTINUE
 6102 CONTINUE
      V(K)=T
      V(K)=T
C     HAVING MODIFIED V BY L-INVERSE...
 6001 CONTINUE
C     NOW THE BACK SOLUTION FOR Y.
      DO 6003 NK=1,N
      K=N+1-NK
      T=V(K)
      KP1=K+1
      IF (KP1 .GT. N) GO TO 6005
      DO 6004 P=KP1,N
      T=T-A(K,P)*Y(P)
 6004 CONTINUE
 6005 CONTINUE
      Y(K)=T/A(K,K)
 6003 CONTINUE
      RETURN
      END


      FORTRAN II FUNCTION AND(X,Y)
B     AND=X*Y
      RETURN
      END


      SUBROUTINE MATOP (A,B,C,L,M,N,LA,MB,NC,IOP)
      DIMENSION A(LA,1),B(MB,1),C(NC,1)
      IF (IOP) 10,40,70
   10 DO 30 I=1,L
      DO 20 J=1,N
      C(I,J)=A(I,J)-B(I,J)
   20 CONTINUE
   30 CONTINUE
   40 DO 60 I=1,L
      DO 50 J=1,N
      C(I,J)=A(I,J)+B(I,J)
   50 CONTINUE
   60 CONTINUE
   70 DO 100 I=1,L
      DO 90 J=1,N
      SUM=0.
      DO 80 II=1,M
      SUM=SUM+A(I,II)*B(II,J)
   80 CONTINUE
      C(I,J)=SUM
   90 CONTINUE
  100 CONTINUE
      RETURN
      END
```

CDCONT and FREADER input:

```
1234567891111111111222222222233333333334444444444555555555566666666667777777778
01234567890123456789012345678901234567890123456789012345678901234567890123456789
```

No. of spectra and No. or possible first neighbors `15  10`

Monomer coefficients in order A C G U, and chainlengths.

```
A4U4    1.      0.      0.      1.      8.
A6U6    1.      0.      0.      1.      12.
A4UAU4  1.      0.      0.      1.      10.
        1.      0.      0.      1.      8.
        1.      0.      0.      1.      10.
        1.      0.      0.      1.      6.
        1.      0.      0.      1.      8.
        1.      0.      0.      1.      10.
        1.      0.      0.      1.      12.
        1.      0.      0.      1.      6.
        1.      0.      0.      1.      8.
        1.      0.      0.      1.      6.
        .5      .5      .5      .5      6.
        .5      .5      .5      .5      6.
A4CG + CGU4  .5      .5      .5      .5      6.
```

Format(10F10.4)

Control card (see SUPSPC description for details), the format is (A8,2X,4(F8.3,2X),2(E13.6,2X)).

title

spectra in 10F8.4 format

Star card precedes next spectrum

etc. until all spectra are read.

```
A4U4 DS
A4U4-4     350.000   190.000    10.000     8.000  1.004000E+00  8.030000E+01
  .0006  -.0314  -.0074  -.0154  -.0234   .0166   .0043   .0080   .CC91   .0082
  .0077   .0010  -.0004  -.0032  -.0031  -.0038   .0004   .0032   .0041   .0035
  .0041   .0036   .0055   .0057   .0088   .0091   .0117   .0143   .0199   .0199
  .0227   .0236   .0235   .0251   .0241   .0238   .0217   .0164   .0136   .0153
  .0200   .0244   .0338   .0414   .0483   .0539   .0595   .0635   .0711   .0781
  .0915   .1100   .1360   .1707   .2189   .2788   .3560   .4554   .5857   .7482
  .9407  1.1610  1.4017  1.6553  1.9156  2.1693  2.4055  2.6222  2.8157  2.9964
 3.1719  3.3517  3.5392  3.7336  3.9423  4.1629  4.3790  4.5831  4.7732  4.9395
 5.0794  5.1880  5.2614  5.2962  5.3028  5.2553  5.1760  5.0388  4.8565  4.6125
 4.3041  3.9107  3.4572  2.9173  2.3231  1.6974  1.0793   .4845  -.0432  -.4970
 -.8367 -1.0693 -1.1939 -1.2193 -1.1699 -1.0668  -.9197  -.7425  -.5600  -.3682
 -.1910  -.0344   .1062   .2430   .3721   .4887   .5764   .6719   .7495   .8348
  .9167  1.0187  1.1256  1.2376  1.3453  1.4398  1.5111  1.5307  1.4522  1.3964
 1.2326   .9940   .6814   .3174  -.0727  -.4415  -.7904 -1.1210 -1.3321 -1.5015
-1.5523 -1.5197 -1.3943 -1.2333 -1.0161  -.7963  -.5569  -.2906  -.0830   .1956
  .5394   .6258   .7984  1.4444  2.0881  1.8642  1.7202  4.3116  5.8872  2.7440
**********************************************************************************
A6U6 DS
A6U6-4.1   350.000   190.000    10.000     5.000  1.151000E+00  7.960000E+01
  .0205   .0205   .0274   .0136   .0205   .0136   .0120   .0075   .0064   .0045
```

```
-1.0949  -.7679  -.5593  -.5188   .2885  -.8366  -.6298  -.9156  -.7818  -.0277
*****************************************************************************
A4CG + CGU4 DS
A4CG-4 C   350.000   195.000    10.000     8.000  9.280001E-01  8.2CCCOOE+C1
  .0235    .0191    .0147    .0102    .0014   -.0030    .0007    .0022    .0010    .0038
  .0060    .0089    .0095    .0080    .0057    .0015   -.0021   -.0068   -.0070   -.0088
 -.0156   -.0186   -.0199   -.0186   -.0197   -.0221   -.0248   -.0273   -.0280   -.0293
 -.0300   -.0336   -.0365   -.0385   -.0387   -.0371   -.0410   -.0427   -.0432   -.0429
 -.0443   -.0474   -.0513   -.0552   -.0554   -.0588   -.0618   -.0621   -.0654   -.0631
 -.0599   -.0559   -.0516   -.0441   -.0350   -.0176    .0130    .0518    .1025    .1669
  .2469    .3470    .4597    .5915    .7346    .8832   1.0293   1.1767   1.3192   1.4566
 1.5807   1.7007   1.8177   1.9354   2.0535   2.1751   2.2859   2.3894   2.4763   2.5474
 2.6026   2.6473   2.6558   2.6439   2.6126   2.5486   2.4702   2.3543   2.2049   2.0173
 1.8014   1.5624   1.3099   1.0425    .7587    .4767    .2032   -.0588   -.2931   -.5145
 -.7072   -.8768  -1.0010  -1.0855  -1.1295  -1.1411  -1.1185  -1.0781  -1.0208   -.9485
 -.8723   -.7943   -.7123   -.6404   -.5678   -.5020   -.4333   -.3655   -.3065   -.2365
 -.1765   -.1181   -.0558    .0217    .1137    .1894    .2848    .3631    .4067    .3969
  .3566    .2721    .0910   -.1333   -.4005   -.6462   -.9590  -1.2359  -1.4833  -1.6774
-1.8407  -1.9337  -2.0201  -2.0445  -2.0714  -2.0113  -1.9067  -1.7891  -1.7326  -1.5272
 -.9794   -.8734   -.6657    .0986   1.3710
*****************************************************************************
6.    .    .   1.    .    .    .    .    .    .    .    .    .    .    .    .
10.   .    .   1.    .    .    .    .    .    .    .    .    .    .    .    .
6.    .    .   2.    .    .    .    .    .   1.    .    .    .    .    .    .
5.   1.   1.    .    .    .    .    .    .    .    .    .    .    .    .    .
7.   1.   1.    .    .    .    .    .    .    .    .    .    .    .    .    .
2.    .   2.    .    .    .    .    .   1.    .    .    .    .    .    .    .
4.    .   2.    .    .    .    .    .   1.    .    .    .    .    .    .    .
6.    .   2.    .    .    .    .    .   1.    .    .    .    .    .    .    .
8.    .   2.    .    .    .    .    .   1.    .    .    .    .    .    .    .
2.   2.    .    .    .    .   1.    .    .    .    .    .    .    .    .    .
4.   2.    .    .    .    .   1.    .    .    .    .    .    .    .    .    .
2.    .    .    .    .    .   1.   2.    .    .    .    .    .    .    .    .
3.    .   1.    .    .   1.    .    .    .    .    .    .    .    .    .    .
2.   1.   1.    .   1.    .    .    .    .    .    .    .    .    .    .    .
3.   1.    .    .    .    .   1.    .    .    .    .    .    .    .    .    .
   1
DSPOLY        6
POLY A-U
1.
      )
POLY C-G
                      1.
POLY AU-AU
              .   .5                        .5
POLY CG-CG
                           .5       .5
POLY AC-GU
     .5             .5
POLY AG-CU
          .5                  .5
```

last spectrum

last star card.
This is the X or Q matrix -- the coefficients of the contributions in the input spectra. They are in the order AA, AC, AG, AU, CA, CC, CG, GA, GC, UA for double strands. The format is 16F5.1. The vertical order is the same as input for monomer coefficients and input spectra

This is the last row of the X or Q matrix. This is input to PREADER, 1 set will be calc'd. This is INDX, which cues the program as to the type of PREADER input. The other options are: DSOLIG, AND SSOLIG. SSPOLY may also equal INDX. 6 is the No. of calc'ns to be made. (A6,4X,I5). After INDX follows Title and the F values. They are the first neighbor frequencies for polymers, and the count of first neigbor contributions of each type for oligomers. The order is again AA, AC, AG,... , Format(16F5.1)

The program then asks for coefficients for the end monomers. This card will be blank for polymer calculations . For oligomer calculations, the input will be the same as the input for the original input spectra... 4 coefficients, followed by the chainlength in 10F10.4 format.

```
1234567891111111111222222222233333333344444444455555555566666666677777777778
0123456789012345678901234567890123456789012345678901234567890123456789012345678890
```

0 0 0 0 3 3 9 4 6 3 7 3

F. Program INTERP

This program takes a set of spectra at a limited number of temperatures and interpolates spectral values for temperatures in between. Consider the interpolation process at one wavelength. The values are calculated from the quadratic equation determined by the three measured points with corresponding temperatures (a triplet) surrounding the interpolation temperature. There are two such temperature triplets for each interpolated point (except for points between the first and second, or last and next-to-last temperatures, where there is only one triplet). The interpolated value is calculated for each triplet and then the two values are averaged with a weighting corresponding to the temperature difference between the central points of the triplets and the interpolation temperature. The weighting factor is 0.5 for each triplet when the interpolation temperature is halfway between two measured points; and 1.0 and 0.0 when an interpolated value equals one of the measured points.

Input is demonstrated for interpolating the temperatures 3, 15, and 25°C from measured 0, 10, 20, and 30° $A_6C_8U_6$ spectra.

The program requires the same unlisted routines as program GETNSM.

0 3 0 0 3 9 0 0 0 7 4


```
      PROGRAM INTERP(INPUT,OUTPUT,PUNCH)
      DIMENSION CD(15,150), T(15),  CON(7,15),ID(8), W(150)
      DIMENSION A(15,3,150), TRIP(3), TI(15)
      DO 5 J=1,15
      DO 5 I=1,150
    5 CD(J,I) = 0.
      READ 970, YMAX,YINC,ISY
    1 READ 902, (ID(I), I=1,8)
      READ 901, NTEMP, L
      READ 903, (T(I), I=1,NTEMP)
      READ 903, (TI(I), I=1,L)
      READ 900, CHK
      READ 915, (CON(I,1), I=1,7)
      XINC = CON(4,1)/10.
      MPTS =(CON(2,1) - CON(3,1))/XINC
      FMPTS = MPTS
      XB = FMPTS/10.
      LB = XB
      XB = XB - FLOAT(LB)
      IF (XB.EQ.0.) GO TO 300
      MB = FLOAT(LB)*10. + 10.
      GO TO 301
  300 MB = MPTS
  301 CONTINUE
      IF(MPTS.GE.100) 11,12
   11 NBEG = MPTS - 100 + 1
      GO TO 13
   12 NBEG = 1
      PRINT 901, NBEG
   13 XMAX = CON(2,1) - FLOAT(NBEG-1)*XINC
      NPTS = MPTS - NBEG - 1
      READ 903, (CD(1,J),J=1,MB)
      READ 900, CHK
      DO 10 I=2,NTEMP
      READ 900, CHK
      READ 915, (CON(J,I), J=1,7)
      READ 903, (CD(I,J), J=1,MB)
   10 READ 900, CHK
      DO 15 J=1,MPTS
   15 W(J) = CON(2,1) - FLOAT(J-1)*XINC
C     GET COEFFS FOR FIT TO ALL CONNECTED TRIPLETS OF T(I) (MEASD TEMPS)
      NN = NTEMP - 2
      DO 30 I=1,NN
      DO 20 K=1,3
   20 TRIP(K) = T(I+K-1)
      ITZERO = I+1
      DO 30 JJ = 1,MPTS
      CALL QSOLV(TRIP, CD(I,JJ), A , ITZERO,JJ)
   30 CONTINUE
C     SEARCH FOR PROPER TRIPLETS OF T(I) MATCHING THE TI(I) (INTERP=0)
      LK = 1
      NNN = NTEMP - 1
      DO 60 I=1,L
      IK = 1
      DO 50 K=LK,NNN
      IK = K
      IF(TI(I).GE.T(K) .AND. TI(I).LT.T(K+1)) GO TO 55
   50 CONTINUE
```

```
C     CALC INTERP CD FROM TRIPLET COEFFS AND TI(I)
   55 D1 = TI(I) - T(K)
      D2 = T(K+1) - TI(I)
      D3 = T(K+1) - T(K)
      JK =IK+1
      LK = IK
      IF(IK.NE.1) GO TO 56
       CD2 =0.    $D1=1.    $D2=1.    $D3=1.    $IK=2
   56 IF(IK.NE.NNN) GO TO 57
       CD2 =0.    $D1=1.    $D2=1.    $D3=1.
   57 CONTINUE
      PRINT 901, IK,JK
      PRINT 903, D1,D2,D3
      CON(1,I) = TI(I)
      DO 58 J=1,MPTS
      CD1 = A(IK,1,J) + A(IK,2,J)*TI(I) + A(IK,3,J)*TI(I)**2
      IF(LK.EQ.1 .OR. IK.EQ.NNN) GO TO 58
      CD2 = A(JK,1,J) + A(JK,2,J)*TI(I) + A(JK,3,J)*TI(I)**2
   58 CD(J) = (D2*CD1 + D1*CD2)/D3
      CALL SMOOTH(MPTS,CD,CD)
      PRINT 910
      PRINT 902,(ID(J), J=1,8)
      PRINT 920, TI(I)
      PRINT 971
      CALL PRNT(W,CD,MPTS)
      PRINT 910
      PRINT 902, (ID(J), J=1,8)
      PRINT 920, TI(I)
      CALL PRNPLT(W(NBEG), CD(NBEG),XMAX, XINC, YMAX, YINC, 0.,ISY,NPTS)
      PUNCH 902, (ID(J), J=1,8)
      PUNCH 916, (CON(J,I), J=1,7)
      PUNCH 903, (CD(J), J=1,MB)
      PUNCH 811
   60 CONTINUE
      IF(CHK.EQ.4HSTOP) STOP
      GO TO 1
  811 FORMAT (80(1H*))
  900 FORMAT(A4)
  901 FORMAT(2I5)
  902 FORMAT(8A10)
  903 FORMAT(10F8.4)
  909 FORMAT(*1 THESE ARE THE COEFFICIENTS FOR THE FIT TO * /8A10/* FOR
     1THE TEMPS* 3F10.3/* OF THE FORM CD = A(1) + A(2)T + A(3)(T SQUARED
     2)*///6X,* WAVELENGTH*, 5X, * A(1)*, 20X,* A(2)*, 20X, * A(3)*//)
  910 FORMAT(1H1)
  911 FORMAT(8X, F5.1, 5X, E12.5, 2(12X, E12.5))
  915 FORMAT(A8,2X,4(F8.3,2X),2(E13.6,2X))
  916 FORMAT(F8.2,2X,4(F8.3,2X),2(E13.6,2X))
  920 FORMAT(* INTERPOLATED FOR TEMP=* F5.1)
  970 FORMAT(2F10.4,5X,I5)
  971 FORMAT(// 3(15X,* LAMBDA*,3X,* PROPERTY*)//)
      END
```

```
      SUBROUTINE QSOLV(TRIP,Y,A,ITZERO,JJ)
      DIMENSION TRIP(3), Y(3), A(15,3,150), G(3,4)
      DO 10 K=1,3
      G(K,2) = 1.
      G(K,3) = TRIP(K)
      G(K,4) = TRIP(K)*TRIP(K)
   10 G(K,1) = Y(K)
      DO 20 L=1,4
      G(2,L) = G(1,L) - G(2,L)
   20 G(3,L) = G(1,L) - G(3,L)
      F = G(2,3)/G(3,3)
      DO 30 L=1,4
   30 G(3,L) = G(2,L) -  F*G(3,L)
      A(ITZERO,3,JJ) = G(3,1)/G(3,4)
      A(ITZERO,2,JJ) =(G(2,1) - A(ITZERO,3,JJ)*G(2,4))/G(2,3)
      A(ITZERO,1,JJ) =(G(1,1) - A(ITZERO,3,JJ)*G(1,4) - A(ITZERO,2,JJ)*G
     1(1,3))/G(1,2)
      RETURN
      END
```

Input for INTERP

```
1234567891111111111222222222233333333334444444444555555555566666666667777777778
01234567890123456789012345678901234567890123456789012345678901234567890123456789
```

```
(2F10.4,5X,I5)
    Plot parameters CDMAX, Scale, and ISY, see input
    illustration for GETNSM.            Name card
(2I5)    Number of meas and number of interp temps.
         Measured temps                 (10F8.4)
         Temps for interpolation
         Meas spectrum identifier
    Control card (see SUPSPC description)
          Format(A8,2X,4(F8.3,2X),2(E13.6,2X))

    Measured spectral values --

              10F8.4 format

                more spectra

                more spectra

            (4 altogether in this
                  example)

          &HSTOP terminates reading.
```

```
 6.        .18              0
A6C8U6
   4    3
0.       10.       20.       30.
3.       15.       25.
A6C8U6 AT  0 DEG IN TMTB
0          340.000   202.500    10.000     -0.        .875000E+00    .900000E+02
  -.0031   -.0031   -.0031   -.0031   -.0031   -.0031   -.0082   -.0082   -.0031   -.0031
   .0021    .0021    .0021    .0021    .0021    .0021   -.0134   -.0082   -.0082    .0021
   .0021    .0021    .0175    .0175    .0175    .0175    .0226    .0278    .0278    .0432
   .0432    .0535    .0638    .0638    .0689    .0792    .0895    .0998    .1203    .1461
   .1975    .2592    .3415    .4392    .5729    .7375    .9278   1.1335   1.3495   1.5861
  1.8381   2.0952   2.3626   2.6352   2.9129   3.1803   3.4478   3.7049   3.9415   4.1781
  4.3941   4.5946   4.7798   4.9341   5.0781   5.1963   5.2683   5.3095   5.3198   5.2786
  5.1963   5.0678   4.9032   4.7026   4.4712   4.2243   3.9415   3.6483   3.3346   3.0209
  2.6969   2.3626   2.0078   1.6426   1.2621    .8763    .4855    .1203   -.2242   -.5174
 -.7642   -.9648  -1.0985  -1.1911  -1.2219  -1.2219  -1.1911  -1.1345  -1.0471   -.9545
 -.8517   -.7642   -.6562   -.5688   -.4865   -.4145   -.3374   -.2808   -.2088   -.1419
 -.0545    .0483    .1615    .2849    .4341    .5781    .7169    .8249    .8763    .8712
  .7838    .6295    .4032    .1306   -.1934   -.5688   -.9648  -1.2939  -1.5459  -1.7002
-1.7157  -1.7259  -1.2837   -.8105   -.2654    .3621   1.0718
**********************************************************************************
A6C8U6 AT 10 DEG IN TMTB
10         340.000   202.500    10.000     -0.        .875000E+00    .900000E+02
  -.0056    .0047    .0047    .0047    .0047    .0047    .0099   -.0004    .0047    .0047
   .0099    .0099    .0099    .0099   -.0004   -.0004   -.0056   -.0004   -.0107   -.0107
```

```
-1.9959  -1.9959  -1.3910  -1.3170  -1.2470  -1.1750  -1.1030
**********************************************************************************
A6C8U6 AT 30 DEG IN TMTB
30         340.000   202.500    10.000     -0.        .875000E+00    .900000E+02
   .0130    .0027    .0027    .0027    .0027    .0027   -.0025   -.0025    .0027    .0027
  -.0025   -.0025   -.0025   -.0025   -.0025   -.0025   -.0076   -.0025   -.0025   -.0025
  -.0025   -.0025    .0027    .0027    .0027    .0027    .0078    .0130    .0130    .0284
   .0284    .0284    .0387    .0387    .0438    .0438    .0644    .0747    .1055    .1312
   .1724    .2238    .2958    .3730    .4758    .5890    .7278    .8821   1.0467   1.2215
  1.4015   1.5867   1.7718   1.9724   2.1678   2.3735   2.5792   2.7850   2.9804   3.1758
  3.3507   3.5101   3.6541   3.7775   3.8701   3.9472   3.9987   4.0192   4.0090   3.9678
  3.8958   3.7878   3.6541   3.4844   3.3044   3.0987   2.8672   2.6255   2.3427   2.0392
  1.7050   1.3707   1.0261    .6815    .3421    .0284   -.2596   -.5013   -.7225   -.9025
 -1.0362  -1.1442  -1.2059  -1.2368  -1.2368  -1.2162  -1.1648  -1.1185  -1.0619  -1.0105
  -.9590   -.9025   -.8459   -.7996   -.7379   -.6865   -.6299   -.5939   -.5425   -.4962
  -.4396   -.3676   -.2956   -.2236   -.1259   -.0128    .0952    .1930    .2855    .3421
  .3472    .3164    .2341    .1158   -.0642   -.3059   -.5888   -.8459  -1.0568  -1.2008
-1.2573  -1.4116  -1.1133   -.8973   -.7225   -.5579   -.4139
$STOP
```

```
1234567891111111111222222222233333333334444444444555555555566666666667777777778
01234567890123456789012345678901234567890123456789012345678901234567890123456789
```

0 0 0 0 3 9 0 0 7 6

G. Program INTERP2

This program performs exactly the same sort of procedure for wavelengths as INTERP does for temperatures. For instance, data may have been recorded at 2.5 nm intervals, but is needed at 1 nm intervals for comparisons with other spectra. The sample input following performs just such a procedure.

Once again the average quadratic fit to connected triplets of measured wavelengths is used. A slight difficulty arises at extrema in the calculated curves since the measured values may not precisely locate their position. This is especially a problem when the CD bands are sharp. Therefore, it is good practice to check the interpolated values at the extrema and correct them if they are inappropriate.

The program requires the same unlisted routines as program GETNSM.

```
      PROGRAM INTERP2(INPUT,OUTPUT,PUNCH)
C
C ADJUST 803 FORMAT FOR APPROPRIATE FORM OF INPUT
C DELETE OR CHANGE STATEMENT BEFORE #15 W(J) = ...* FOR PROPER SCALING.
C
      DIMENSION CD(150), T(150), CON(7)   ,ID(8), W(150)
      DIMENSION A(150,3), TRIP(3), TI(150)
      EQUIVALENCE (T,W)
      DO 5 I=1,15)
    5 CD(I) = 0.
      READ 970, YMAX,YINC,ISY
      READ 903, XINT
    1 READ 902, (ID(I), I=1,8)
      READ 915, (CON(I)  , I=1,7)
      XINC = CON(4)  /10.
      CON(4) = XINT*10.
      MPTS =(CON(2)   - CON(3)  )/XINC
      LPTS =(CON(2)   - CON(3)  )/XINT
      IF(LPTS.GE.100) 11,12
   11 NBEG = LPTS - 100 + 1
      GO TO 13
   12 NBEG = 1
      PRINT 901, NBEG
   13 XMAX = CON(2)   - FLOAT(NBEG-1)*XINT
      NPTS = LPTS - NBEG - 1
C
C ADJUST 803 FORMAT TO APPROPRIATE FORM FOR INPUT
C
      READ 803, (CD(J),J=1,MPTS)
      READ 900, CHK
      DO 15 J=1,MPTS
C
C DELETE OR CHANGE NEXT CARD FOR APPROPRIATE SCALING
C
   15 W(J) = CON(2)   - FLOAT(J-1)*XINC
      DO 16 J=1,LPTS
   16 TI(J)= CON(2) - FLOAT(J-1)*XINT
C     GET COEFFS FOR FIT TO ALL CONNECTED TRIPLETS OF T(I) (MEASD TEMPS)
      PRINT 910
      PRINT 939, (ID(KK), KK=1,8),XINC, XINT
      NN  = MPTS - 2
      DO 30 I=1,NN
      DO 20 K=1,3
   20 TRIP(K) = T(I+K-1)
      ITZERO = I+1
      CALL QSOLV2(TRIP, CD(I),    A , ITZERO)
   30 PRINT 940, (TRIP(KK), KK=1,3), (A(ITZERO,KK),KK=1,3)
C     SEARCH FOR PROPER TRIPLETS OF T(I) MATCHING THE TI(I) (INTERP#D)
      NNN = MPTS - 1
      LK = 1
      DO 60 I=1,LPTS
      DO 50 K=LK,NNN
      IK = K
      IF(TI(I).LE.T(K) .AND. TI(I).GT.T(K+1)) GO TO 55
   50 CONTINUE
C     CALC INTERP CD FROM TRIPLET COEFFS AND TI(I)
   55 D1 = TI(I) - T(K)
      D2 = T(K+1) - TI(I)
```

```
      D2 = T(K+1) - T(K)
      JK =IK+1
      LK = IK
      IF(IK.NE.1) GO TO 56
       CD2 =0.    $D1=1.    $D2=1.    $D3=1.    $IK=2
   56 IF(IK.NE.NNN) GO TO 57
       CD2 =0.    $D1=1.    $D2=1.    $D3=1.
   57 CONTINUE
      CD1 = A(IK,1)    + A(IK,2)  *TI(I) + A(IK,3)  *TI(I)**2
      IF(LK.EQ.1 .OR. IK.EQ.NNN) GO TO 58
      CD2 = A(JK,1)    + A(JK,2)  *TI(I) + A(JK,3)  *TI(I)**2
   58 CD(I) = (D2*CD1 + D1*CD2)/D3
   60 CONTINUE
      CALL SMOOTH(LPTS,CD,CD)
      PRINT 910
      PRINT 902,(ID(J), J=1,8)
      PRINT 920, XINC, XINT
      PRINT 971
      CALL PRNT(TI,CD,LPTS)
      PRINT 910
      PRINT 902, (ID(J), J=1,8)
      PRINT 920, XINC, XINT
      CALL PRNPLT(TI(NBEG),CD(NBEG),XMAX, XINT, YMAX, YINC, 0.,ISV,NPTS)
      PUNCH 902, (ID(J), J=1,8)
      PUNCH 915, (CON(J)  , J=1,7)
      PUNCH 903, (CD(J), J=1,LPTS)
      PUNCH 811
      IF(CHK.EQ.4HSTOP) STOP
      GO TO 1
  811 FORMAT (80(1H*))
  900 FORMAT(A4)
  901 FORMAT(3I5)
  902 FORMAT(8A10)
  903 FORMAT(6X, 4E16.7)
  905 FORMAT(10F8.4)
  910 FORMAT(1H1)
  915 FORMAT(A8,2X,4(F8.3,2X),2(E13.6,2X))
  916 FORMAT(F8.2,2X,4(F8.3,2X),2(E13.6,2X))
  920 FORMAT(* INTERPOLATED FROM WAVELENGTH INTERVAL =*F8.2* NM TO INT =
     1*F8.2* NM*)
  939 FORMAT(*0 THESE ARE THE COEFFICIENTS FOR THE FIT TO * /8A10/* INPU
     1T WAVELENTH INTERVAL =*F8.2* NM, INTERPOLATED INT =*F8.2* NM*//
     2  15X,* WAVELENGTHS*, 40X,* COEFFICIENTS*)
  940 FORMAT(5X, 3(F8.2,5X), 15X, 3(E12.5,5X))
  970 FORMAT(2F10.4,5X,I5)
  971 FORMAT(// 3(15X,* LAMBDA*,3X,* PROPERTY*)//)
      END


      SUBROUTINE DSOLV2(TRIP,Y,A,ITZERO)
      DIMENSION TRIP(3), Y(3), A(150,3)     , G(3,4)
      DO 10 K=1,3
      G(K,2) = 1.
      G(K,3) = TRIP(K)
      G(K,4) = TRIP(K)*TRIP(K)
   10 G(K,1) = Y(K)
      DO 20 L=1,4
      G(2,L) = G(1,L) - G(2,L)
   20 G(3,L) = G(1,L) - G(3,L)
      F = G(2,3)/G(3,3)
      DO 30 L=1,4
   30 G(3,L) = G(2,L) - F*G(3,L)
      A(ITZERO,3)     = G(3,1)/G(3,4)
      A(ITZERO,2)     =(G(2,1) - A(ITZERO,3)   *G(2,4))/G(2,3)
      A(ITZERO,1)     =(G(1,1) - A(ITZERO,3)   *G(1,4) - A(ITZERO,2)   *G
     1(1,3))/G(1,2)
      RETURN
      END
```

```
Input to INTERP2                          12345678911111111112222222222333333333344444444445555555555666666666677777777778
                                                  01234567890123456789012345678901234567890123456789012345678901234567890

Plot parameters, see INTERP               6.        .2                  0
Desired wavelength interval (F8.4)          1.
title                                      POLY A-U                                  MEAS BY DANA CARROLL
Control card (A8,2X,3(F8.3,2X)            POLY A-U   310.       210.        25.
SPectra (First card has all zero values).
                                                                                .10E            .20E
                                                .45E              .80E            1.20E           1.65E
                                                 2.05E            2.45E           2.80E           3.20E
                                                3.45E            3.75E            3.80E           3.65E
format is 6X,4E16.7 but can be changed to suit   3.25E            2.60E            1.60E            .45E
data.  (803 FORMAT)                             -.85E           -1.80E           -2.00E          -1.75E
                                               -1.25E           -.75E            -.35E           -.10E
                                                .10E            .30E             .55E            .85E
                                               1.00E            .75E             0.              -1.25E
                                          ****************************************************************************
                                           POLY C-G
                                          POLY C-G  310.       210.        25.

                                               -.05E            -.05E            -.05E           .10E
                                               .25E             .35E             .45E            .55E
another spectrum                               .70E             1.00E            1.30E           1.55E
                                               1.70E            1.65E            1.45E           1.25E
                                               1.25E            1.30E            1.30E           1.30E
                                               1.25E            1.15E            .85E            .50E
                                               .15E             -.15E            -.45E           -.65E
                                               -.75E            -.80E            -.90E           -1.00E
                                               -1.20E           -1.40E           -1.60E          -1.75E
4HSTOP in place of last asterisk card     STOP
terminates execution.

                                          12345678911111111112222222222333333333344444444445555555555666666666677777777778
                                                  01234567890123456789012345678901234567890123456789012345678901234567890
```

0 0 0 0 3 9 0 0 7 8

H. Program DIFFSPC

This program performs the subtraction:

$$AFACT*\underset{\sim}{A} - BFACT*\underset{\sim}{B}$$

$\underset{\sim}{A}$ and $\underset{\sim}{B}$ are spectra and AFACT and BFACT are scalar multipliers. The spectra must begin at the same wavelength and also end at identical wavelengths. There are four ways data can be handled in the program: Normal, "NEWFACT", "READB", and "ZEROB". The program controller "CHK" determines which of the options is executed.

Normal input: The CHK card can be anything, usually it is the spectrum name. The identifier for the output follows the second spectrum, then comes AFACT and BFACT.

"NEWFACT" input: When CHK = 7HNEWFACT, a new identifier and scalar factors are read. The most recently read A and B spectra are used.

"READB" input: When CHK = 5HREADB; a new B spectrum, identifier, and multipliers are read. The most recently entered A spectrum is used.

"ZEROB" input: When CHK = 5HZEROB; an A spectrum, identifier, and multipliers are read. The B array is set to zero. This input is used when one simply wants to generate a line printer plot for a spectrum.

Unlisted routines required:

PRNT (see SMOOSP listing)

PRNPLT (see SMOOSP listing)

PLSCAL (see SMOOSP listing)

```
      PROGRAM DIFSPC(INPUT, OUTPUT)
      DIMENSION A(150), B(150), ID(8), DIFF(150), W(150), CON(7)
      EQUIVALENCE (CON(2), WMAX)
      READ 908, YMAX, YINCR,ISY
    1 READ 900, CHK
      IF (CHK.EQ. 4HSTOP) STOP
      IF (CHK.EQ. 7HNEWFACT) GO TO 6
      IF (CHK.EQ. 5HREADB) GO TO 5
      IF (CHK.NE.5HZEROB) GO TO 4
      READ 900, DUM
      DO 3 I=1,150
    3 B(I) = 0.
    4 CONTINUE
      READ 915, (CON(I), I=1,7)
      XINC = CON(4)/10.
      MPTS =(CON(2) - CON(3))/XINC
      FMPTS = MPTS
      XB = FMPTS/10.
      LB = XB
      XB = XB - FLOAT(LB)
      IF (XB.EQ.0.) GO TO 300
      MB = FLOAT(LB)*10. + 10.
      GO TO 301
  300 MB = MPTS
  301 CONTINUE
      READ 902, (A(I), I=1,MB)
      READ 900, DUM
      IF (CHK.EQ.5HZEROB) GO TO 6
    5 READ 900, DUM
      READ 900, DUM
      READ 902, (B(I), I=1,MB)
      READ 900, DUM
    6 READ 904, (ID(I), I=1,8)
      READ 903, AFACT, BFACT
      DO 11 I=1,MPTS
   11 DIFF(I) = AFACT*A(I) - BFACT*B(I)
      DO 10 I=1,MPTS
   10 W(I) = WMAX - FLOAT(I-1)*XINC
      PRINT 905, (ID(I), I=1,8)
      PRINT 906
      CALL PRNT(W, DIFF, MPTS)
      NBEG = (WMAX - 305.)/XINC +1.
      NPTS = MPTS - NBEG + 1
      PRINT 905, (ID(I), I=1,8)
      CALL PRNPLT(W(NBEG), DIFF(NBEG), 305.,1.,YMAX,YINCR,0,ISY,NPTS)
      GO TO 1
  900 FORMAT(A7)
  901 FORMAT(I5, 5X, 2F10.5)
  902 FORMAT(10F8.4)
  903 FORMAT(2F10.5)
  904 FORMAT(8A10)
  905 FORMAT(*1THIS IS THE DIFFERENCE SPECTRUM FOR* /8A10///)
  906 FORMAT(3(15X,* LAMBDA*, 3X, * ELLIPTICITY*)//)
  907 FORMAT(1H1)
  908 FORMAT(2F10.5, 9X, I1)
  915 FORMAT(A8,2X,4(F8.3,2X),2(E13.6,2X))
      END
```

Input to DIFFSPC

```
1234567891111111111222222222233333333334444444444555555555566666666667777777778
01234567890123456789012345678901234567890123456789012345678901234567890123456789
```

Plot card, (2F10.5,9X,I1)
CHK card -- normal input
Control card (A8,2X,4(F8.3,2X),2(E13.6,2X)

```
 6.        .18                0
A6C8U6 AT  0 DEG IN TMTB
0          340.000   202.500    10.000    -0.       .875000E+00    .900000E+02
```

Spectral points for A
(10F8.4)

```
 -.0031  -.0031  -.0031  -.0031  -.0031  -.0031  -.0082  -.0082  -.0031  -.0031
  .0021   .0021   .0021   .0021   .0021   .0021  -.0134  -.0082  -.0082   .0021
  .0021   .0021   .0175   .0175   .0175   .0175   .0226   .0278   .0278   .0432
  .0432   .0535   .0638   .0638   .0689   .0792   .0895   .0998   .1203   .1461
  .1975   .2592   .3415   .4392   .5729   .7375   .9278  1.1335  1.3495  1.5861
 1.8381  2.0952  2.3626  2.6352  2.9129  3.1803  3.4478  3.7049  3.9415  4.1781
 4.3941  4.5946  4.7798  4.9341  5.0781  5.1963  5.2683  5.3095  5.3198  5.2786
 5.1963  5.0678  4.9032  4.7026  4.4712  4.2243  3.9415  3.6483  3.3346  3.0209
 2.6969  2.3626  2.0078  1.6426  1.2621   .8763   .4855   .1203  -.2242  -.5174
 -.7642  -.9648 -1.0985 -1.1911 -1.2219 -1.2219 -1.1911 -1.1345 -1.0471  -.9545
 -.8517  -.7642  -.6562  -.5688  -.4865  -.4145  -.3374  -.2808  -.2088  -.1419
 -.0545   .0483   .1615   .2849   .4341   .5781   .7169   .8249   .8763   .8712
  .7838   .6295   .4032   .1306  -.1934  -.5688  -.9648 -1.2939 -1.5459 -1.7002
-1.7157 -1.7259 -1.2837  -.8105  -.2654   .3621  1.0718
```

Star card

```
**********************************************************************************
A6C8U6 AT 10 DEG IN TMTB
10         340.000   202.500    10.000    -0.       .875000E+00    .900000E+02
```

Points for B

```
 -.0056   .0047   .0047   .0047   .0047   .0047   .0099  -.0004   .0047   .0047
  .0099   .0099   .0099   .0099  -.0004  -.0056  -.0004  -.0107  -.0107
 -.0107  -.0107  -.0056  -.0056  -.0158  -.0158  -.0107  -.0056  -.0056   .0099
  .0099   .0202   .0304   .0304   .0356   .0356   .0459   .0562   .0767   .1024
  .1539   .2156   .2876   .3853   .5087   .6527   .8327  1.0282  1.2236  1.4396
 1.6607  1.8973  2.1442  2.3859  2.6224  2.8693  3.1162  3.3527  3.5893  3.8156
 4.0316  4.2424  4.4379  4.6230  4.7670  4.8956  4.9676  5.0087  5.0087  4.9676
 4.8956  4.7670  4.6230  4.4430  4.2219  3.9853  3.7230  3.4299  3.1367  2.8230
 2.4887  2.1544  1.7996  1.4242  1.0539   .6682   .2979  -.0570  -.3810  -.6638
 -.8901 -1.0701 -1.1833 -1.2656 -1.2861 -1.2861 -1.2450 -1.1987 -1.1216 -1.0496
 -.9673  -.8901  -.7924  -.7153  -.6330  -.5507  -.4633  -.3964  -.3141  -.2473
 -.1804  -.0981  -.0056   .0973   .2259   .3596   .5087   .6476   .7504   .8173
  .8122   .7299   .5653   .3133  -.0004  -.3656  -.7513 -1.1010 -1.3838 -1.5793
-1.6564 -1.6770 -1.4301 -1.1318  -.7821  -.3398   .1642
```

Title for plot
APACT and BPACT for preceding A and B
CHK card -- normal

```
**********************************************************************************
A6C8U6(0 DEG) - A6C8U6(10 DEG)
1.        1.
A6C8U6 AT 20 DEG IN TMTB
20         340.000   202.500    10.000    -0.       .875000E+00    .900000E+02
```

new A spectrum

```
 -.0179  -.0076  -.0076  -.0076  -.0076  -.0076  -.0025  -.0025   .0027   .0027
  .0078  -.0025  -.0025  -.0025   .0078   .0078   .0027   .0078   .0078   .0078
  .0078  -.0025   .0027   .0027   .0027   .0027   .0078   .0130   .0130   .0181
  .0181   .0181   .0181   .0284   .0335   .0438   .0541   .0644   .0850   .1107
  .1518   .2032   .2855   .3730   .4964   .6404   .7998   .9747  1.1598  1.3552
 1.5661  1.7821  1.9981  2.2192  2.4558  2.6924  2.9187  3.1552  3.3712  3.5770
 3.7827  3.9627  4.1272  4.2712  4.3947  4.4821  4.5438  4.5644  4.5541  4.5027
 4.4204  4.3021  4.1478  3.9678  3.7570  3.5307  3.2684  2.9855  2.6821  2.3581
 2.0135  1.6484  1.2832   .9181   .5591   .1930  -.1362  -.4396  -.7019  -.9230
-1.0876 -1.2162 -1.2882 -1.3190 -1.3088 -1.2882 -1.2368 -1.1802 -1.1133 -1.0516
 -.9796  -.9230  -.8459  -.7790  -.7070  -.6350  -.5682  -.5116  -.4396  -.3830
 -.3059  -.2236  -.1413  -.0385   .0695   .1930   .3112   .4192   .5118   .5684
  .5632   .4912   .3370   .1055  -.1773  -.4910  -.8048 -1.0825 -1.2933 -1.4270
-1.4939 -1.4939 -1.3910 -1.3190 -1.2470 -1.1750 -1.1030
**********************************************************************************
A6C8U6 AT 30 DEG IN TMTB
30         340.000   202.500    10.000    -0.       .875000E+00    .900000E+02
   .0130   .0027   .0027   .0027   .0027   .0027  -.0025  -.0025   .0027   .0027
  -.0025  -.0025  -.0025  -.0025  -.0025  -.0025  -.0076  -.0025  -.0025  -.0025
  -.0025  -.0025   .0027   .0027   .0027   .0027   .0078   .0130   .0130   .0284
```

```
new B spectrum                                  .0284   .0284   .0387   .0387   .0438   .0438   .0644   .0747   .1055   .1312
                                                .1724   .2238   .2958   .3730   .4758   .5890   .7278   .8821  1.0467  1.2215
                                               1.4015  1.5867  1.7718  1.9724  2.1678  2.3735  2.5792  2.7850  2.9804  3.1758
                                               3.3507  3.5101  3.6541  3.7775  3.8701  3.9472  3.9987  4.0192  4.0090  3.9678
                                               3.8958  3.7878  3.6541  3.4844  3.3044  3.0987  2.8672  2.6255  2.3427  2.0392
                                               1.7050  1.3707  1.0261   .6815   .3421   .0284  -.2596  -.5013  -.7225  -.9025
                                              -1.0362 -1.1442 -1.2059 -1.2368 -1.2368 -1.2162 -1.1648 -1.1185 -1.0619 -1.0105
                                               -.9590  -.9025  -.8459  -.7996  -.7379  -.6865  -.6299  -.5939  -.5425  -.4962
                                               -.4396  -.3676  -.2956  -.2236  -.1259  -.0128   .0952   .1930   .2855   .3421
                                                .3472   .3164   .2341   .1158  -.0642  -.3059  -.5888  -.8459 -1.0568 -1.2008
                                              -1.2573 -1.4116 -1.1133  -.8973  -.7225  -.5579  -.4139
                                              ******************************************************************************
Title for plot                                2*A6C8U6(20 DEG) - A6C8U6(30 DEG)
AFACT AND BFACT for preceding A and B           2.          1.
CHK = NEWFACT, use old A and B, with new factors.  NEWFACT
Title                                         A6C8U6(20 DEG) - A6C8U6(30 DEG)
new AFACT and BFACT                           1.          1.
CHK = READB, use old A, read new B            READB
                                              A6C8U6 AT 40 DEG IN TMTB
                                              40          340.000    202.500     10.000     -0.        .875000E+00     .900000E+02
                                               -.0109  -.0109  -.0109  -.0109  -.0109  -.0109  -.0058  -.0058  -.0006  -.0006
new B spectrum                                  .0045   .0045   .0045   .0045   .0045   .0045  -.0006   .0045   .0045   .0045
                                                .0045   .0045   .0097   .0097   .0097   .0097   .0148   .0200   .0200   .0251
                                                .0251   .0251   .0251   .0354   .0405   .0405   .0611   .0714   .0920   .1177
                                                .1588   .2102   .2720   .3491   .4417   .5548   .6834   .8274   .9714  1.1257
                                               1.2748  1.4394  1.6040  1.7531  1.9177  2.0925  2.2674  2.4525  2.6377  2.8125
                                               2.9771  3.1262  3.2497  3.3628  3.4348  3.4914  3.5222  3.5222  3.4914  3.4400
                                               3.3577  3.2497  3.1160  2.9565  2.7765  2.5914  2.3908  2.1594  1.9177  1.6657
                                               1.3828  1.1000   .8171   .5240   .2462  -.0263  -.2732  -.5046  -.7155  -.8955
                                              -1.0292 -1.1269 -1.1886 -1.2195 -1.2092 -1.1886 -1.1372 -1.0909 -1.0138  -.9623
                                               -.9109  -.8646  -.8080  -.7720  -.7206  -.6795  -.6332  -.5869  -.5458  -.4995
                                               -.4532  -.3915  -.3298  -.2578  -.1600  -.0572   .0611   .1588   .2514   .3182
                                                .3440   .3440   .2925   .1742   .0148  -.1755  -.3863  -.5920  -.7823  -.9366
                                              -1.0446 -1.0858 -1.0549  -.9623  -.8080  -.5818  -.2938
                                              ******************************************************************************
Title for plot.                               A6C8U6(20 DEG) - A6C8U6(40 DEG)
New AFACT and BFACT                           1.          1.
CHK = ZEROB -- set B = all zeroes.            ZEROB
                                              A6C8U6 AT 50 DEG IN TMTB
                                              50          340.000    202.500     10.000     -0.        .875000E+00     .900000E+02
                                               -.0109  -.0109  -.0109  -.0109  -.0109  -.0109  -.0058  -.0058  -.0006  -.0006
                                                .0045   .0045   .0045   .0045   .0045   .0045  -.0006   .0045   .0045   .0045
Read new A                                      .0045   .0045   .0097   .0097   .0097   .0097   .0148   .0200   .0200   .0251
                                                .0251   .0251   .0251   .0251   .0200   .0200   .0302   .0405   .0611   .0868
                                                .1280   .1794   .2411   .3080   .4005   .5034   .6217   .7554   .8891  1.0331
                                               1.1822  1.3262  1.4805  1.6194  1.7737  1.9177  2.0720  2.2160  2.3600  2.4937
                                               2.6068  2.7251  2.8177  2.8794  2.9308  2.9565  2.9565  2.9462  2.9154  2.8640
                                               2.7920  2.6942  2.5811  2.4422  2.2931  2.1182  1.9382  1.7274  1.5165  1.2851
                                               1.0537   .8120   .5600   .3080   .0714  -.1498  -.3452  -.5252  -.6640  -.7720
                                               -.8543  -.9109  -.9315  -.9520  -.9520  -.9623  -.9418  -.9160  -.8800  -.8595
                                               -.8286  -.8132  -.7772  -.7618  -.7412  -.7103  -.6743  -.6486  -.6075  -.5612
                                               -.5252  -.4738  -.4223  -.3709  -.2938  -.2012  -.1035  -.0058   .0971   .1742
                                                .2102   .2000   .1485   .0714  -.0469  -.1960  -.3658  -.5406  -.6795  -.7926
                                               -.8903 -1.0446  -.9932  -.9932 -1.0240 -1.0652 -1.1269
                                              ******************************************************************************
Title for plot                                A6C8U6(50 DEG)
new AFACT                                     1.
CHK = 4HSTOP, terminates the program.         STOP

                                              123456789111111111122222222223333333333444444444455555555556666666666777777777778
                                                      012345678901234567890123456789012345678901234567890123456789012345678900
```

0 0 0 0 3 9 0 0 7 8 0

REFERENCES FOR APPENDIX ON COMPUTER PROGRAMS

1. Blum, A.D., Ph.D. thesis, Univ. of Calif., Berkeley (1971).

2. Tomlinson, B.L., Ph.D. thesis, Univ. of Calif., Berkeley (1968).

APPENDIX 2:

DOUBLE STRAND FIRST NEIGHBOR CONTRIBUTIONS

The spectra for the double strand first neighbor contributions were presented in section IV.A.2. This appendix gives the values digitized at 1 nm intervals from 350 to 196 nm.

DS AA CONTRIBUTION

WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY
320.0	-.1388	278.0	8.0678	236.0	-1.2534
319.0	-.1401	277.0	8.5116	235.0	-.9401
318.0	-.1307	276.0	8.9354	234.0	-.6396
317.0	-.1360	275.0	9.3437	233.0	-.3928
316.0	-.1433	274.0	9.7105	232.0	-.1889
315.0	-.1361	273.0	10.0371	231.0	-.0364
314.0	-.1170	272.0	10.3120	230.0	.1101
313.0	-.1028	271.0	10.5541	229.0	.2675
312.0	-.0914	270.0	10.7126	228.0	.4777
311.0	-.0829	269.0	10.8214	227.0	.6933
310.0	-.0714	268.0	10.8639	226.0	.9193
309.0	-.0727	267.0	10.8690	225.0	1.1627
308.0	-.0791	266.0	10.7264	224.0	1.4097
307.0	-.0776	265.0	10.5548	223.0	1.5992
306.0	-.0830	264.0	10.3489	222.0	1.7286
305.0	-.0845	263.0	10.1208	221.0	1.7232
304.0	-.0787	262.0	9.8065	220.0	1.4606
303.0	-.0619	261.0	9.4234	219.0	1.1606
302.0	-.0436	260.0	8.9530	218.0	.5129
301.0	-.0300	259.0	8.3593	217.0	-.3509
300.0	-.0212	258.0	7.5601	216.0	-1.2781
299.0	.0259	257.0	6.5397	215.0	-2.3572
298.0	.1062	256.0	5.3215	214.0	-3.4225
297.0	.2038	255.0	3.9243	213.0	-4.1317
296.0	.3518	254.0	2.3304	212.0	-4.8868
295.0	.5636	253.0	.7236	211.0	-5.3359
294.0	.8344	252.0	-.8316	210.0	-5.4921
293.0	1.1649	251.0	-2.2637	209.0	-5.4183
292.0	1.5456	250.0	-3.5132	208.0	-5.1260
291.0	1.9749	249.0	-4.4918	207.0	-4.5850
290.0	2.4363	248.0	-5.1683	206.0	-4.2715
289.0	2.9429	247.0	-5.5638	205.0	-3.9480
288.0	3.4677	246.0	-5.6880	204.0	-3.0567
287.0	4.0119	245.0	-5.5893	203.0	-2.6432
286.0	4.5331	244.0	-5.2490	202.0	-2.4045
285.0	5.0321	243.0	-4.8057	201.0	-2.5146
284.0	5.5041	242.0	-4.2546	200.0	-2.8253
283.0	5.9712	241.0	-3.7018	199.0	-2.2066
282.0	6.4158	240.0	-3.1218	198.0	-1.8326
281.0	6.8294	239.0	-2.5732	197.0	-3.0680
280.0	7.2437	238.0	-2.0672	196.0	-3.7328
279.0	7.6590	237.0	-1.6258	0.0	0.0000

DS AC CONTRIBUTION

WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY
320.0	-.5756	278.0	16.2217	236.0	1.7471
319.0	-.5822	277.0	17.4735	235.0	1.8481
318.0	-.5824	276.0	18.7328	234.0	2.0379
317.0	-.6552	275.0	19.9663	233.0	2.2504
316.0	-.7674	274.0	21.2449	232.0	2.7174
315.0	-.8954	273.0	22.3452	231.0	3.0965
314.0	-.9838	272.0	23.3770	230.0	3.6849
313.0	-1.1268	271.0	24.2367	229.0	4.3736
312.0	-1.2890	270.0	24.7473	228.0	5.1887
311.0	-1.4746	269.0	24.8296	227.0	5.8556
310.0	-1.6737	268.0	24.7417	226.0	6.2709
309.0	-1.8795	267.0	24.1545	225.0	6.6871
308.0	-2.0900	266.0	23.2376	224.0	6.6977
307.0	-2.2779	265.0	22.0587	223.0	6.2502
306.0	-2.5771	264.0	20.5130	222.0	5.5477
305.0	-2.8418	263.0	18.8369	221.0	4.7154
304.0	-3.1597	262.0	17.0719	220.0	3.3204
303.0	-3.5465	261.0	15.1953	219.0	1.4434
302.0	-3.9515	260.0	13.2698	218.0	-.3057
301.0	-4.4194	259.0	11.1833	217.0	-2.4422
~~300.0~~	~~-4.8919~~	~~258.0~~	~~9.1496~~	~~216.0~~	~~-4.3404~~
299.0	-5.2576	257.0	7.1541	215.0	-6.9425
298.0	-5.5458	256.0	5.4140	214.0	-8.6470
297.0	-5.6759	255.0	3.8102	213.0	-10.5641
296.0	-5.5185	254.0	2.5808	212.0	-12.1009
295.0	-5.0696	253.0	1.6854	211.0	-13.6644
294.0	-4.4176	252.0	1.2454	210.0	-14.8058
293.0	-3.3497	251.0	1.2261	209.0	-15.7100
292.0	-2.0419	250.0	1.5081	208.0	-15.7380
291.0	-.4463	249.0	2.0888	207.0	-15.6330
290.0	1.3435	248.0	2.6056	206.0	-14.4178
289.0	3.2595	247.0	3.0798	205.0	-14.3642
288.0	5.2188	246.0	3.5557	204.0	-13.1422
287.0	6.9769	245.0	3.7505	203.0	-11.5338
286.0	8.5421	244.0	3.7278	202.0	-8.3106
285.0	9.8590	243.0	3.5249	201.0	-7.1316
284.0	10.9894	242.0	3.2989	200.0	-7.9492
283.0	11.9531	241.0	2.8924	199.0	.0282
282.0	12.7140	240.0	2.5651	198.0	1.7848
281.0	13.4349	239.0	2.1979	197.0	-3.3288
280.0	14.2809	238.0	1.8827	196.0	-14.2848
279.0	15.1934	237.0	1.7994	0.0	0.0000

DS AG CONTRIBUTION

WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY
320.0	1.1154	278.0	-6.6021	236.0	.5449
319.0	1.0242	277.0	-7.1051	235.0	.0788
318.0	.8989	276.0	-7.4277	234.0	-.6108
317.0	.9544	275.0	-7.6332	233.0	-.9434
316.0	1.0430	274.0	-7.7453	232.0	-1.2643
315.0	1.0730	273.0	-7.7670	231.0	-1.1550
314.0	1.0623	272.0	-7.7846	230.0	-1.2676
313.0	1.0737	271.0	-7.7846	229.0	-1.2261
312.0	1.1855	270.0	-7.4435	228.0	-1.7293
311.0	1.2627	269.0	-7.0956	227.0	-2.0267
310.0	1.2766	268.0	-6.7458	226.0	-2.1233
309.0	1.4123	267.0	-6.3363	225.0	-2.3780
308.0	1.5071	266.0	-5.2448	224.0	-2.6938
307.0	1.5809	265.0	-4.4747	223.0	-2.8861
306.0	1.8151	264.0	-3.8348	222.0	-3.4350
305.0	1.9668	263.0	-3.6110	221.0	-3.8948
304.0	2.1763	262.0	-3.4852	220.0	-3.7128
303.0	2.4248	261.0	-3.6105	219.0	-4.3284
302.0	2.7370	260.0	-3.9655	218.0	-4.4976
301.0	3.1339	259.0	-4.4124	217.0	-4.0847
300.0	3.6089	258.0	-4.8112	216.0	-4.6205
299.0	3.8455	257.0	-4.9712	215.0	-3.4506
298.0	4.0016	256.0	-4.7838	214.0	-3.0875
297.0	4.1023	255.0	-4.0503	213.0	-3.8382
296.0	3.9723	254.0	-2.5259	212.0	-3.1859
295.0	3.5247	253.0	-.9623	211.0	-3.0008
294.0	2.8222	252.0	.5589	210.0	-3.7127
293.0	1.7332	251.0	2.2132	209.0	-3.8639
292.0	.5166	250.0	3.9199	208.0	-4.0408
291.0	-.8354	249.0	5.2531	207.0	-4.0827
290.0	-2.0926	248.0	6.3987	206.0	-2.7583
289.0	-3.3956	247.0	7.0948	205.0	.3088
288.0	-4.5724	246.0	7.2116	204.0	-.5104
287.0	-5.4813	245.0	7.0921	203.0	1.0015
286.0	-5.9596	244.0	6.5851	202.0	2.9720
285.0	-6.1765	243.0	6.0133	201.0	9.6768
284.0	-6.1824	242.0	5.0656	200.0	18.8279
283.0	-6.2681	241.0	4.4114	199.0	13.5233
282.0	-6.1929	240.0	3.3631	198.0	15.4403
281.0	-6.1476	239.0	2.5895	197.0	32.0442
280.0	-6.2894	238.0	1.7913	196.0	50.6369
279.0	-6.4652	237.0	1.0171	0.0	0.0000

DS AU CONTRIBUTION

WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY
320.0	.6593	278.0	6.9501	236.0	9.5919
319.0	.6667	277.0	7.4659	235.0	9.7357
318.0	.5630	276.0	8.3977	234.0	9.5275
317.0	.5793	275.0	9.5529	233.0	9.5652
316.0	.5919	274.0	10.9526	232.0	9.6227
315.0	.5553	273.0	12.3933	231.0	10.0676
314.0	.4114	272.0	13.9231	230.0	10.5453
313.0	.2482	271.0	15.2142	229.0	11.0648
312.0	.1235	270.0	16.5254	228.0	11.3278
311.0	.0429	269.0	17.6026	227.0	11.6729
310.0	-.0175	268.0	18.4712	226.0	12.0867
309.0	.0015	267.0	18.8547	225.0	12.3605
308.0	.1079	266.0	19.6920	224.0	12.2345
307.0	.1426	265.0	19.8864	223.0	11.7826
306.0	.2198	264.0	19.6485	222.0	10.8840
305.0	.2626	263.0	18.8033	221.0	9.7256
304.0	.2685	262.0	17.6877	220.0	9.3058
303.0	.2124	261.0	16.2024	219.0	7.6168
302.0	.1916	260.0	14.1601	218.0	7.1721
301.0	.2024	259.0	11.7125	217.0	7.0419
300.0	.3712	258.0	9.3107	216.0	6.4005
299.0	.3661	257.0	6.7371	215.0	6.9320
298.0	.3061	256.0	4.7532	214.0	6.9011
297.0	.2741	255.0	3.4701	213.0	4.7378
296.0	.1234	254.0	3.4245	212.0	4.6479
295.0	-.2410	253.0	3.7467	211.0	3.1945
294.0	-.6815	252.0	4.4782	210.0	.9039
293.0	-1.1638	251.0	5.6342	209.0	-.3344
292.0	-1.4276	250.0	7.1401	208.0	-1.0704
291.0	-1.5182	249.0	8.6388	207.0	-2.5802
290.0	-1.2704	248.0	9.8112	206.0	-1.0047
289.0	-.8483	247.0	10.8476	205.0	.7576
288.0	-.1117	246.0	11.3800	204.0	-1.9640
287.0	.7169	245.0	11.5247	203.0	-.8314
286.0	1.8684	244.0	11.4877	202.0	1.7910
285.0	2.9314	243.0	11.5675	201.0	6.7275
284.0	3.9469	242.0	11.2622	200.0	16.1753
283.0	4.6606	241.0	11.2470	199.0	15.5244
282.0	5.1566	240.0	10.8916	198.0	18.7655
281.0	5.5740	239.0	10.4634	197.0	35.0735
280.0	5.9711	238.0	9.9925	196.0	43.4087
279.0	6.3358	237.0	9.6675	0.0	0.0000

DS CA CONTRIBUTION

WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY
320.0	-.5298	278.0	-4.4247	236.0	-11.7543
319.0	-.4509	277.0	-4.5107	235.0	-11.2980
318.0	-.3779	276.0	-4.9170	234.0	-10.6581
317.0	-.3873	275.0	-5.4756	233.0	-10.1625
316.0	-.3874	274.0	-6.1248	232.0	-9.9819
315.0	-.3471	273.0	-6.7686	231.0	-10.0332
314.0	-.3209	272.0	-7.4082	230.0	-9.9838
313.0	-.2393	271.0	-7.9100	229.0	-10.0933
312.0	-.2478	270.0	-8.5390	228.0	-9.9719
311.0	-.1851	269.0	-8.8377	227.0	-9.8132
310.0	-.0122	268.0	-9.1788	226.0	-9.5107
309.0	.0278	267.0	-9.3997	225.0	-9.1152
308.0	.1537	266.0	-9.9540	224.0	-8.2496
307.0	.2659	265.0	-10.4031	223.0	-6.9287
306.0	.3291	264.0	-10.7181	222.0	-4.9041
305.0	.4113	263.0	-10.8471	221.0	-2.9088
304.0	.4991	262.0	-10.9388	220.0	-.9807
303.0	.5742	261.0	-11.0622	219.0	1.6981
302.0	.5857	260.0	-10.9601	218.0	4.0953
301.0	.5935	259.0	-10.5752	217.0	5.7273
300.0	.5068	258.0	-10.1664	216.0	7.7087
299.0	.5679	257.0	-9.6240	215.0	8.2254
298.0	.5881	256.0	-9.3382	214.0	8.0262
297.0	.4388	255.0	-9.3092	213.0	8.2259
296.0	.2386	254.0	-9.7007	212.0	7.2252
295.0	-.0528	253.0	-10.1560	211.0	6.3291
294.0	-.3280	252.0	-10.5182	210.0	5.3546
293.0	-.6361	251.0	-11.3655	209.0	3.3972
292.0	-.9770	250.0	-12.3921	208.0	1.1876
291.0	-1.3914	249.0	-13.4698	207.0	-1.5919
290.0	-1.9872	248.0	-14.4852	206.0	-4.7021
289.0	-2.5249	247.0	-15.2849	205.0	-7.9586
288.0	-3.0797	246.0	-15.7999	204.0	-8.1676
287.0	-3.5839	245.0	-15.9991	203.0	-9.6129
286.0	-4.1057	244.0	-15.9929	202.0	-12.5901
285.0	-4.4799	243.0	-15.8322	201.0	-15.2387
284.0	-4.7827	242.0	-15.3784	200.0	-18.3668
283.0	-4.7819	241.0	-14.9025	199.0	-18.3530
282.0	-4.7266	240.0	-14.1210	198.0	-20.2715
281.0	-4.5588	239.0	-13.5484	197.0	-28.8057
280.0	-4.3692	238.0	-12.7707	196.0	-25.4253
279.0	-4.2930	237.0	-12.1166	0.0	0.0000

DS CC CONTRIBUTION

WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY
320.0	-.7387	278.0	12.6283	236.0	.3279
319.0	-.5822	277.0	13.9145	235.0	.5394
318.0	-.4168	276.0	15.0528	234.0	.9066
317.0	-.4709	275.0	16.0982	233.0	1.1662
316.0	-.5062	274.0	17.0171	232.0	1.5476
315.0	-.5412	273.0	17.9222	231.0	1.5813
314.0	-.5661	272.0	18.5951	230.0	2.0317
313.0	-.6345	271.0	19.1622	229.0	2.3456
312.0	-.7722	270.0	19.2380	228.0	3.1654
311.0	-.9155	269.0	19.2453	227.0	3.9258
310.0	-.9376	268.0	19.0366	226.0	4.4643
309.0	-1.0746	267.0	18.4729	225.0	5.0564
308.0	-1.1539	266.0	17.1852	224.0	5.5560
307.0	-1.2396	265.0	15.8152	223.0	5.7167
306.0	-1.4462	264.0	14.4556	222.0	5.9325
305.0	-1.6595	263.0	13.2194	221.0	6.1950
304.0	-1.8918	262.0	11.8954	220.0	5.7103
303.0	-2.2516	261.0	10.6205	219.0	5.1093
302.0	-2.6766	260.0	9.4659	218.0	4.6255
301.0	-3.1671	259.0	8.4188	217.0	3.3336
300.0	-3.7611	258.0	7.6564	216.0	2.5812
299.0	-4.2296	257.0	6.9479	215.0	.1132
298.0	-4.6128	256.0	6.2187	214.0	-2.0643
297.0	-4.9765	255.0	5.4471	213.0	-3.7240
296.0	-5.1325	254.0	4.4577	212.0	-6.2421
295.0	-4.9888	253.0	3.7034	211.0	-8.3954
294.0	-4.5633	252.0	3.2731	210.0	-10.1163
293.0	-3.7035	251.0	2.8078	209.0	-11.7881
292.0	-2.5785	250.0	2.3157	208.0	-13.4300
291.0	-1.1894	249.0	1.9705	207.0	-15.6927
290.0	.3160	248.0	1.5100	206.0	-18.2227
289.0	2.0097	247.0	1.1359	205.0	-21.3004
288.0	3.6370	246.0	1.0939	204.0	-20.3555
287.0	5.0858	245.0	.9560	203.0	-21.4555
286.0	6.2008	244.0	.7887	202.0	-20.5026
285.0	7.1150	243.0	.5708	201.0	-22.2641
284.0	7.8162	242.0	.5151	200.0	-24.1099
283.0	8.6028	241.0	.2466	199.0	-15.0825
282.0	9.1363	240.0	.2923	198.0	-12.6330
281.0	9.7976	239.0	.1855	197.0	-26.2841
280.0	10.6744	238.0	.2357	196.0	-37.1994
279.0	11.6498	237.0	.3626	0.0	0.0000

DS CG CONTRIBUTION

WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY
320.0	.6319	278.0	-19.2662	236.0	-4.4407
319.0	.5994	277.0	-20.4706	235.0	-4.6943
318.0	.5366	276.0	-21.6150	234.0	-4.9652
317.0	.6011	275.0	-22.6410	233.0	-5.1103
316.0	.7330	274.0	-23.7114	232.0	-5.4881
315.0	.8584	273.0	-24.5644	231.0	-5.4926
314.0	.8429	272.0	-25.3861	230.0	-5.8079
313.0	.9229	271.0	-26.1204	229.0	-6.2745
312.0	1.0448	270.0	-26.4395	228.0	-6.9781
311.0	1.2085	269.0	-26.3012	227.0	-7.3633
310.0	1.3562	268.0	-26.2210	226.0	-7.3525
309.0	1.5272	267.0	-25.7675	225.0	-7.5860
308.0	1.7099	266.0	-24.7684	224.0	-7.1848
307.0	1.8477	265.0	-23.8079	223.0	-6.3537
306.0	2.1636	264.0	-22.5462	222.0	-5.5006
305.0	2.3942	263.0	-21.5346	221.0	-4.7564
304.0	2.6582	262.0	-20.5753	220.0	-3.0509
303.0	2.9913	261.0	-19.7534	219.0	-1.2889
302.0	3.3015	260.0	-18.9551	218.0	.2213
301.0	3.7551	259.0	-17.9013	217.0	2.2432
300.0	4.1970	258.0	-16.4430	216.0	3.6991
299.0	4.5051	257.0	-14.5361	215.0	6.5548
298.0	4.5987	256.0	-12.4859	214.0	7.6893
297.0	4.5185	255.0	-10.0130	213.0	8.4127
296.0	4.0163	254.0	-7.2238	212.0	9.2609
295.0	3.1266	253.0	-4.5951	211.0	9.8715
294.0	2.0017	252.0	-2.2494	210.0	9.1937
293.0	.3966	251.0	-.5455	209.0	8.7603
292.0	-1.4695	250.0	.6475	208.0	6.8748
291.0	-3.6089	249.0	1.0019	207.0	4.8541
290.0	-5.8562	248.0	1.0563	206.0	2.3755
289.0	-8.1279	247.0	.7841	205.0	2.0726
288.0	-10.3494	246.0	.1802	204.0	-.5681
287.0	-12.2012	245.0	-.4244	203.0	-2.0057
286.0	-13.6564	244.0	-1.1276	202.0	-5.2672
285.0	-14.7309	243.0	-1.7480	201.0	-3.6509
284.0	-15.5681	242.0	-2.4678	200.0	4.6723
283.0	-16.2077	241.0	-2.8353	199.0	-3.8891
282.0	-16.6345	240.0	-3.3208	198.0	-4.2754
281.0	-16.9874	239.0	-3.6552	197.0	13.7153
280.0	-17.6252	238.0	-3.8697	196.0	41.9344
279.0	-18.3800	237.0	-4.2464	0.0	0.0000

DS GA CONTRIBUTION

WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY
320.0	-.1658	278.0	11.1231	236.0	3.1752
319.0	-.1584	277.0	11.7483	235.0	3.3200
318.0	-.1394	276.0	12.3297	234.0	3.4606
317.0	-.1580	275.0	12.8587	233.0	3.6068
316.0	-.2382	274.0	13.3632	232.0	3.8742
315.0	-.3093	273.0	13.7255	231.0	4.0628
314.0	-.3056	272.0	14.0055	230.0	4.3284
313.0	-.3688	271.0	14.1907	229.0	4.7444
312.0	-.4328	270.0	14.1537	228.0	5.2522
311.0	-.4920	269.0	13.8444	227.0	5.6130
310.0	-.5335	268.0	13.5553	226.0	5.8340
309.0	-.5841	267.0	13.0221	225.0	6.1962
308.0	-.6372	266.0	12.2841	224.0	6.1796
307.0	-.6881	265.0	11.5167	223.0	5.9022
306.0	-.8298	264.0	10.5773	222.0	5.5657
305.0	-.9314	263.0	9.6965	221.0	5.2621
304.0	-1.0757	262.0	8.8287	220.0	4.5403
303.0	-1.2662	261.0	8.0000	219.0	3.5794
302.0	-1.4684	260.0	7.2169	218.0	2.7614
301.0	-1.7582	259.0	6.3685	217.0	1.4096
300.0	-2.0168	258.0	5.4805	216.0	.2671
299.0	-2.2277	257.0	4.5490	215.0	-1.9009
298.0	-2.3620	256.0	3.7417	214.0	-3.4164
297.0	-2.4185	255.0	2.8641	213.0	-5.3111
296.0	-2.2984	254.0	2.0439	212.0	-6.9230
295.0	-1.9865	253.0	1.3750	211.0	-8.8907
294.0	-1.5631	252.0	.9069	210.0	-10.5971
293.0	-.8421	251.0	.6881	209.0	-12.3659
292.0	.0457	250.0	.6626	208.0	-12.9076
291.0	1.1151	249.0	.9297	207.0	-13.0466
290.0	2.2835	248.0	1.2584	206.0	-12.0879
289.0	3.5117	247.0	1.6294	205.0	-12.5335
288.0	4.7767	246.0	2.0600	204.0	-11.9226
287.0	5.8935	245.0	2.4071	203.0	-11.3563
286.0	6.8499	244.0	2.6585	202.0	-8.8891
285.0	7.6404	243.0	2.8667	201.0	-7.7997
284.0	8.2839	242.0	2.9809	200.0	-10.1987
283.0	8.8409	241.0	2.9963	199.0	-5.2748
282.0	9.2395	240.0	3.0677	198.0	-4.9667
281.0	9.6276	239.0	3.0610	197.0	-12.0456
280.0	10.0891	238.0	2.9932	196.0	-23.0454
279.0	10.5888	237.0	3.0942	0.0	0.0000

OS GC CONTRIBUTION

WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY
320.0	-1.9780	278.0	31.5704	236.0	-1.0789
319.0	-1.8798	277.0	34.6384	235.0	-.6789
318.0	-1.6964	276.0	37.5069	234.0	.2069
317.0	-1.8397	275.0	40.2980	233.0	.6976
316.0	-2.0534	274.0	42.9248	232.0	1.4276
315.0	-2.2226	273.0	45.3333	231.0	1.5322
314.0	-2.3655	272.0	47.5865	230.0	2.1596
313.0	-2.5686	271.0	49.5162	229.0	2.5244
312.0	-2.9694	270.0	50.6370	228.0	3.9148
311.0	-3.2685	269.0	51.3222	227.0	4.8895
310.0	-3.4421	268.0	51.5915	226.0	5.3270
309.0	-3.8358	267.0	51.4202	225.0	5.7128
308.0	-4.1338	266.0	49.2200	224.0	5.9627
307.0	-4.4361	265.0	46.9911	223.0	5.5677
306.0	-5.0006	264.0	44.3230	222.0	5.3925
305.0	-5.3970	263.0	41.6227	221.0	4.4955
304.0	-5.9232	262.0	38.6406	220.0	2.1122
303.0	-6.5756	261.0	35.3618	219.0	.1946
302.0	-7.3061	260.0	32.1098	218.0	-2.8219
301.0	-8.1795	259.0	28.5957	217.0	-7.5753
300.0	-9.1378	258.0	25.2957	216.0	-10.9200
299.0	-9.6741	257.0	21.8229	215.0	-17.7042
298.0	-10.3812	256.0	18.2488	214.0	-22.6629
297.0	-10.2688	255.0	14.2723	213.0	-26.1105
296.0	-10.0275	254.0	9.4905	212.0	-30.8705
295.0	-9.1779	253.0	5.5508	211.0	-34.1186
294.0	-7.7955	252.0	2.4984	210.0	-35.2803
293.0	-5.5323	251.0	-.2464	209.0	-35.4148
292.0	-2.8804	250.0	-2.5445	208.0	-34.1038
291.0	.2810	249.0	-4.1539	207.0	-31.1256
290.0	3.4820	248.0	-5.3816	206.0	-28.1814
289.0	7.0098	247.0	-6.0019	205.0	-26.8130
288.0	10.5505	246.0	-5.7351	204.0	-17.3110
287.0	13.7611	245.0	-5.4318	203.0	-8.9393
286.0	16.3035	244.0	-4.9428	202.0	-1.3093
285.0	18.5098	243.0	-4.5444	201.0	-.8553
284.0	20.2517	242.0	-3.7391	200.0	-5.1245
283.0	22.1013	241.0	-3.5421	199.0	14.8442
282.0	23.5823	240.0	-2.7242	198.0	20.6710
281.0	25.2427	239.0	-2.3919	197.0	-5.8690
280.0	27.2059	238.0	-1.9613	196.0	-28.8651
279.0	29.3922	237.0	-1.3586	0.0	0.0000

DS UA CONTRIBUTION

WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY	WAVELENGTH	PROPERTY
320.0	-.5000	278.0	-7.5144	236.0	-2.5644
319.0	-.5886	277.0	-7.6652	235.0	-2.8228
318.0	-.3576	276.0	-8.3269	234.0	-2.8187
317.0	-.4068	275.0	-9.1999	233.0	-2.8076
316.0	-.3258	274.0	-10.1344	232.0	-2.7430
315.0	-.2162	273.0	-10.9170	231.0	-3.5057
314.0	.1114	272.0	-11.9430	230.0	-4.4803
313.0	.4666	271.0	-12.5048	229.0	-5.3127
312.0	.7153	270.0	-12.9866	228.0	-5.7283
311.0	.9317	269.0	-13.0906	227.0	-6.4253
310.0	.9639	268.0	-12.8470	226.0	-7.1968
309.0	.8948	267.0	-11.8882	225.0	-8.0519
308.0	.6143	266.0	-11.4120	224.0	-8.5047
307.0	.5902	265.0	-10.2412	223.0	-8.5654
306.0	.5062	264.0	-8.7312	222.0	-8.1775
305.0	.3804	263.0	-6.6349	221.0	-7.3144
304.0	.4444	262.0	-4.3760	220.0	-7.9217
303.0	.5075	261.0	-2.2860	219.0	-6.9437
302.0	.5773	260.0	.1783	218.0	-7.9159
301.0	.5220	259.0	2.7532	217.0	-8.0337
300.0	.2924	258.0	4.9068	216.0	-7.2120
299.0	.2522	257.0	7.2367	215.0	-7.9820
298.0	.1951	256.0	7.9606	214.0	-7.7421
297.0	.1480	255.0	7.4361	213.0	-4.2509
296.0	.1337	254.0	5.2652	212.0	-2.7746
295.0	.4243	253.0	2.4397	211.0	1.6885
294.0	.7164	252.0	-.5569	210.0	6.5527
293.0	1.0889	251.0	-3.6695	209.0	10.2003
292.0	1.2547	250.0	-6.6554	208.0	13.0728
291.0	1.1599	249.0	-9.1189	207.0	15.6326
290.0	.6390	248.0	-10.3879	206.0	14.2125
289.0	-.0023	247.0	-11.1469	205.0	14.3168
288.0	-.8734	246.0	-10.7593	204.0	18.4622
287.0	-1.7606	245.0	-9.5334	203.0	15.3302
286.0	-3.0013	244.0	-8.4774	202.0	9.9469
285.0	-4.0421	243.0	-7.7092	201.0	2.2888
284.0	-5.1199	242.0	-6.8021	200.0	-15.3127
283.0	-5.8018	241.0	-6.4234	199.0	-13.6709
282.0	-6.2617	240.0	-5.4364	198.0	-10.6835
281.0	-6.5748	239.0	-4.4582	197.0	-28.2908
280.0	-6.8577	238.0	-3.4530	196.0	-47.3007
279.0	-7.0306	237.0	-2.7335	0.0	0.0000

LEGAL NOTICE

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

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