Lawrence Berkeley National Laboratory
Recent Work

**Title**
POLYNUCLEOTIDE CIRCULAR DICHROISM CALCULATIONS: USE OF ALL ORDER POLARIZABILITY THEORY

**Permalink**
https://escholarship.org/uc/item/34k0b2q4

**Author**
Cech, Carol Martinson.

**Publication Date**
1975-05-01
POLYNUCLEOTIDE CIRCULAR DICHROISM CALCULATIONS:
USE OF ALL ORDER POLARIZABILITY THEORY

Carol Martinson Cech
(Ph. D. thesis)

May 1975

Prepared for the U. S. Energy Research and Development Administration under Contract W-7405-ENG-48

For Reference

Not to be taken from this room
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.
This thesis is dedicated to my sister, Lisa, in memory
Polynucleotide Circular Dichroism Calculations:
Use of All Order Polarizability Theory

Abstract
Carol Martinson Cech

Polarizability theory, an all order classical coupled oscillator theory in which monomer absorption bands shapes are explicitly taken into account, is applied to the calculation of polynucleotide circular dichroism (CD) and absorption. CD calculations are shown for ApA and oligo adenylic acid of varying chain length, in both RNA and B-DNA geometry, and the advantages of this theory are assessed. By introducing an effective dielectric constant, polarizability theory predicts CD spectra in good agreement with measured spectra. Variations in monomer parameters are tested in an attempt to eliminate remaining discrepancies between calculated and measured spectra.

Optical property calculations are presented for poly A:U, poly AU:AU, poly G:C, and poly GC:GC in RNA, B-DNA, and C-DNA conformations. The calculated CD spectra are found to be sensitive to both geometry and sequence. Agreement with the measured CD spectra of poly rA:rU, poly rG:rC, and poly dG:dC is good. Calculations for other sequences and geometries are less satisfactory and are particularly poor for poly
GC:GC in RNA geometry and poly A:U in B-DNA geometry. Attempts to improve agreement with measured spectra by varying monomer properties have been only partially successful for these calculations, but they illustrate the types of changes that may prove to be necessary. Speculation that some sequences may adopt unusual conformations in solution is also discussed. In addition, the dependence of the calculated CD on chain length is examined. Results show that non-nearest neighbor interactions can be important when runs of 3 or more identical base pairs appear in a given sequence.

Calculations are done for the first time for polynucleotides in several non-standard geometries. Results of calculations for poly I in 3 or 4 strand arrays of parallel right hand helices strongly support X-ray evidence that poly I forms such a structure, with bases slightly tilted with respect to the helix axis, at high salt concentration in aqueous solution. Calculations using recently published X-ray coordinates for certain deoxy polynucleotides of simple sequence, some of which are quite different from B-DNA coordinates, are also shown.

Finally, calculations for the CD of all 16 dimers and of double strand oligomers of more complicated sequence are presented. It is concluded that polarizability theory is probably sufficiently accurate and
reliable, but that insufficient understanding of the monomer properties or of the solution conformations of the polynucleotides being studied continues to cause errors in some of the calculations.
ACKNOWLEDGEMENTS

Undertaking and completing this work would not have been possible without the invaluable guidance, support, and occasional prodding given by my research director, Nacho Tinoco. I am also much indebted to Werner Hug for greatly facilitating the initiation of this work by providing the original program ROTOP and many helpful discussions.

I would like to thank many of the third floor Hildebrand people who were there when I arrived, especially Phil Borer, Arlene Blum, Mark Levine, Olke Uhlenbeck, Mike Botchan, and Burt Dorman, for helping me in many ways to get started in research. Life as a graduate student would not have been the same without David Koh and Barbara Dengler, who have cheerfully provided technical assistance and good food during all the years this work progressed. Barbara Dengler is particularly responsible for making sure research was not placed above achievement of a weekly quota of laps at the pool.

More recent members of the group whose scientific advice has been much appreciated are Doug Turner and my "roommate", Ken Breslauer. Alan Levin's help and sympathy with both theoretical and computational problems has been immeasurable. Thanks also go to Soo Freier for answers to any and all quantum mechan-
ical problems and for backup support in non-scientific
discussions with W. Hug. David Gee is thanked for
his careful measurements of oligo and poly A spectra
when they were needed in a hurry. The friendship of
these and other members of the group has made time
spent in the lab most enjoyable, and last minute
proof reading by many of them was especially
appreciated.

Lastly, I am deeply grateful to Tom Cech for his
thorough support; for his unselfish willingness to
help whenever and however needed, and for doing my
share of the house chores the last few weeks this
thesis was in preparation.
Table of Contents

INTRODUCTION ................................................. xi

1. THEORY
   I. Background .............................................. 1
   II. First Order Perturbation Theory ..................... 3
   III. Matrix Method ......................................... 5
   IV. Time Dependent Hartree Theory ....................... 10
   V. Polarizability Theory ................................... 11
   VI. The Two-state Dimer: A Comparison of Theories .......... 15
   VII. Computational Considerations ....................... 18
   VIII. A Note on Intrinsic Monomer CD .................... 19

2. MONOMER INPUT PARAMETERS
   I. Number and Location of Monomer Transitions .......... 21
   II. Polarizabilities ....................................... 23
   III. Transition Dipole Moments ................................ 30
   IV. Monopoles ............................................... 32
   V. Geometry ................................................ 34

3. ApA AND OLIGO A_N CALCULATIONS
   I. Objective ............................................... 40
   II. ApA Calculations ........................................ 40
   III. Oligo A_N Calculations .................................. 47
   IV. Dependence of Calculations on Effective Dielectric Constant ......... 52
   V. Variation of Monomer Parameters ....................... 55
   A. Number of Oscillators .................................... 55
B. Variations in $\theta$ \hspace{1cm} 58
C. Adjusting the Monopoles \hspace{1cm} 59
D. Shifting the Monomer Absorption Bands \hspace{1cm} 61
VI. Discussion \hspace{1cm} 62

4. CALCULATIONS FOR DOUBLE STRANDED POLYNUCLEOTIDES
OF REPEATING SEQUENCE

I. Objective \hspace{1cm} 73
II. RNA Calculations \hspace{1cm} 74
III. B-DNA Calculations \hspace{1cm} 82
IV. C-DNA Calculations \hspace{1cm} 88
V. Parallel and Perpendicular Components \hspace{1cm} 89
VI. Effect of Dielectric Constant \hspace{1cm} 101
VII. Calculations Done with Non-standard
Geometries \hspace{1cm} 105
VIII. Empirical Search for Better Monomer
Parameters \hspace{1cm} 109
A. Monopoles \hspace{1cm} 109
B. Transition Angles \hspace{1cm} 111
C. Shifting the Monomer Absorption Bands \hspace{1cm} 118
IX. Discussion \hspace{1cm} 119a

5. CHAIN LENGTH DEPENDENCE AND THE NEAREST
NEIGHBOR APPROXIMATION \hspace{1cm} 130
I. Objective \hspace{1cm} 130
II. Results \hspace{1cm} 133
III. Discussion \hspace{1cm} 142
6. POLY I AND POLY G CALCULATIONS: SINGLE AND MULTI-STRANDS

I. Objective ........................................ 147
II. Monomer Properties ............................... 150
III. Poly I Calculations ............................. 153
IV. Poly G Calculations .............................. 159
V. Poly I in Other Geometries ...................... 165
VI. What Causes the Long Wavelength Negative CD Band? ................................. 167
VII. Discussion ......................................... 175

7. MISCELLANEOUS RESULTS AND SUGGESTIONS

I. Random Sequence Calculations .................... 179
II. Other Double Strand Sequences .................. 185
III. Dimer Calculations .............................. 186
IV. General Suggestions .............................. 202

APPENDIX: Computer Programs ....................... 206
INTRODUCTION

Molecules that do not have superimposable mirror images have long been known to rotate a beam of plane polarized light. Such molecules, termed optically active, also absorb left and right circularly polarized light to different extents. Circular dichroism (CD) spectroscopy, the measurement of this differential absorption as a function of frequency, has been widely used for monitoring changes in the solution structure of polynucleotides. In comparison to other spectroscopic methods that have been applied to nucleic acid structural studies, e.g. fluorescence, NMR, and Raman scattering, CD is singularly sensitive to helical geometry. Without the ordered helical arrangement of the mononucleotide units, essentially no polymer CD would be observed, since the mononucleotides themselves have virtually no intrinsic optical activity.

CD has been useful for such purposes as verifying the double stranded nature of many naturally occurring forms of RNA and for monitoring transitions between the structural families of DNA. The CD spectrum of double stranded DNA, for instance, has been found to be highly sensitive to variations in temperature, solvent, and ionic strength, and to the presence of certain proteins. Unfortunately, the interpretation of these spectral changes is seldom indisputable, and the usefulness of
CD in actually defining structural details remains frustratingly limited. What is still lacking is a workable theory and a reliable set of monomer properties capable of linking structure and sequence with measured spectra. Theory must be capable of reproducing with reasonable accuracy the observed spectra of molecules whose geometries are already known before it can with any credibility provide insight into, for instance, aspects of DNA structure that might be involved in protein - nucleic acid recognition or other forms of cellular regulation.

The work presented in this thesis is a new attempt to establish a more dependable method of calculating polynucleotide CD and absorption spectra. Earlier attempts to calculate the optical properties of polynucleotides, discussed in Chapter 1, were only partially successful. A number of factors, however, prompted a belief that more accurate calculations could now be done. CNDO-CI molecular orbital calculations that take all valence electrons into account have recently been published for the bases. New studies of the transition moment orientations in some of the bases have also appeared in the literature. CD measurements have now been made on a number of synthetic double stranded polynucleotides of simple repeating sequences, providing an excellent opportunity for assessing in detail the accuracy of the calculations. Finally, theoretical
methods are available, but have not yet been tested on polynucleotides, that are more refined than the approach most commonly applied to polynucleotide CD calculations to date.

In this work, the use of an all order coupled oscillator polarizability theory developed by DeVoe is applied to calculations of polynucleotide CD for the first time. We will refer to this method simply as polarizability theory, since the appearance of complex, frequency dependent monomer polarizabilities in the polymer equations is its most distinguishing feature. [The term polarizability theory as used here should not be confused with Kirkwood's polarizability theory (ref. 26, Chapter 1), a first order approximation that deals with transitions far removed from the frequency range being considered.] A small modification of the theory, the use of an effective dielectric constant, will be evaluated. In addition, variations in the number and properties of monomer absorption bands will be evaluated, in order to gain some understanding of the types of parameter adjustments that may or may not be capable of correcting errors in the calculated spectra. Some calculations will also be done, for purposes of comparison with polarizability theory, with a quantum mechanical matrix method, also all order in interaction energies but lacking the incorporation of monomer absorption band shapes.
Chapter 1

THEORY

I. Background

Clear and concise discussions of the origin of optical activity, from both classical and quantum mechanical points of view, can be found in refs. (1-4). A thorough but more difficult treatment is presented in ref. (5). It is sufficient to begin here with the simple relationship derived quantum mechanically by Rosenfeld in 1928. 6

$$ R_{OA} = \text{Im} \frac{\mu_{OA}}{\omega} \cdot m_{AO} $$  \hspace{1cm} (1)

The rotational strength, $R_{OA}$, describes the intensity of the circular dichroism arising from a transition from molecular state 0 to state A. Im means "imaginary part of", and $\mu_{OA}$ and $m_{AO}$ are the molecular electric dipole and magnetic dipole transition moments, respectively. Following the procedure used in most treatments of polymer optical activity, the transition from state 0 to state A will be referred to hereon as transition K. Only transitions from the ground state to singly excited states will be considered. The absorption for any molecule, optically active or not, can be described in a similar manner by the oscillator strength $D_K$ where
Both the CD and absorption can therefore be readily calculated once $\sim K$ and $\sim K'$ for all transitions in the molecule are known. Unfortunately, molecules as large as polynucleotides are still beyond the scope of currently available *a priori* quantum mechanical methods. Polynucleotide optical properties have been and continue to be calculated by perturbation techniques in which the electronic properties of a mononucleotide unit are assumed to be only minimally altered by transitions occurring in the remaining monomer units of the polymer. On a quantum mechanical level, no electron exchange or electron transfer, i.e. no overlap of molecular orbitals between monomer ground or excited states is allowed. Presumably this is a valid assumption for polynucleotides. In all perturbation approaches, the unperturbed properties of mononucleotides provide the starting point for calculating polymer data. Aside from geometrical coordinates for the polymer, all input parameters, i.e. all monomer properties, are theoretically measurable. In practice, however, many monomer properties must be estimated or theoretically calculated.

It should be noted that magnetically allowed monomer transitions are not required for polynucleotide CD. In polynucleotides, $\sim K$ appears to come largely from the

\[
D_K = \mu K' \mu K
\]
helical arrangement of the strong electrically allowed transitions of the mononucleotide units. In other words, both \( m_K \) and \( m_K \) come from coupling of the monomer transition dipoles \( \mu_i \). The mononucleotides themselves have only a very small intrinsic CD in the near UV caused by perturbations of the base transitions by the sugar-phosphate backbone.

Absorption will be presented throughout this thesis in terms of the extinction coefficient \( \epsilon \), in \( \text{cm}^{-1} \text{ mole}^{-1} \text{ cm}^{-1} \), versus wavelength. Similarly CD will be given as \( \epsilon_L - \epsilon_R \), the difference in extinction coefficients for left and right circularly polarized light. Both calculated and measured data will be presented on a "per nucleotide unit" basis, unless otherwise indicated.

II. First Order Perturbation Theory

Calculations of polynucleotide CD have been carried out in the past with a first order perturbation theory derived from standard time independent perturbation techniques and a Taylor series expansion around 260 nm. A detailed discussion of the full perturbation approach to nucleic acid optical properties can be found in ref. (8). Johnson and Tinoco greatly simplified the approach by introducing the Taylor expansion and truncating it after two terms. Only the CD near 260 is calculated, and it is the sum
of two terms. One term comes from the coupling of the near UV (above 220 nm) transitions of a given base with the polarizability arising from all other transitions on other bases. It has an absorption-like band shape centered at 260 nm. The second term comes from exciton-type coupling of the near UV bands and has a band shape that is the first derivative of the first CD term. Without the Taylor expansion simplification, the second term would generally contain terms to second or higher order in the perturbation energy $V_{ij}$, since degenerate perturbation theory must be used for coupling of identical transitions. The expansion and subsequent truncation brings the entire CD back to first order in $V_{ij}$. Hence the theory will be referred to here as first order perturbation theory.

Despite its simplicity, this technique has been surprisingly successful in predicting the CD above 220 nm of naturally occurring RNA and of DNA under conditions thought to favor the B and C forms.\textsuperscript{7,9-12} Exact shape and position of the observed CD bands are poorly predicted, but the general differences between the calculated spectra for the three forms correspond to experimental observations. It is possible, however, that the success of the simple treatment occurs because random sequences of nucleic acids give rise to a very large number of polymer rotational bands which extend over a wide frequency range but tend to cancel,
allowing the resultant spectra to be roughly approximated by only two terms. This situation is certainly not expected to hold for polymers of simple repeating sequences, and indeed, attempts to calculate CD for poly A<sup>7</sup> and for synthetic double stranded polynucleotides<sup>7,10-12</sup> have been much less successful. Poor results are especially likely in a polymer such as poly GC:GC where 3 strong bands at distinctly separate frequencies between 250 and 280 nm interact. Moreover, it is difficult to make conclusive statements about random sequence calculations when agreement may be occurring somewhat fortuitously.

III. Matrix method

Perturbation theory extended to all order in V<sub>ij</sub> has been successfully applied to CD calculations of polypeptides.<sup>13,14</sup> In this method, the time independent perturbation Hamiltonian H = H<sub>0</sub> + V<sub>ij</sub>, where H<sub>0</sub> is the sum of the unperturbed monomer Hamiltonians and V<sub>ij</sub> is the interaction energy between all transitions i and j on different bases, is written in matrix form as [V<sub>i</sub>δ<sub>ij</sub> + V<sub>ij</sub>], where V<sub>i</sub> is the frequency of transition i and δ<sub>ij</sub> is the Kronecker delta. The matrix is then diagonalized to give the eigenvalues, V<sub>K</sub>, and eigenvector coefficients, C<sub>iK</sub>, of the polymer transitions. Polymer oscillator strengths and rotational strengths are then calculated by eqs. (1) and (2)
using
\[ \mu_K = \sum_i C_{iK} \mu_i \]  
\[ \text{Im} \ m_K = \pi \sum_i C_{iK} \mathbf{R}_i \times \mu_i \] 

where \( \mathbf{R}_i \) is the position vector of \( \mu_i \). This method so far lacks a commonly accepted name and will be referred to here as the matrix method.

The matrix method is similar to first order perturbation theory in that some arbitrary choice of band shape must be assigned to each \( R_K \) or \( D_K \) in order to calculate spectral properties as a function of frequency. In this case, the optical properties are given by

\[ \varepsilon(\nu) = \sum_K D_K f(\nu, \nu_K) \]
\[ \varepsilon_L - \varepsilon_R(\nu) = 4 \sum_K R_K f(\nu, \nu_K) \] 

The methodology is quite different, though, in that a limited basis set for the Hamiltonian is chosen consisting of individual transitions, including transitions representative of the far UV polarizability of each base, if desired. The location of each polymer CD band directly depends on the specific frequency of each monomer band from which it is derived. In particular, each monomer band above 220 nm is coupled independently to all others such that frequency separation between near UV bands is taken into account. We have used this theory occasionally for certain polynucleotide CD calculations in order to look at rotational
strength patterns, since individual polymer transitions do not appear directly in polarizability theory (discussed in next section). In these calculations, gaussian band shapes of arbitrary halfwidth $\Gamma$ are used for $f(\tilde{\nu}, \tilde{\nu}_K)$.

$$f(\tilde{\nu}, \tilde{\nu}_K) = \frac{51.15 \tilde{\nu}_K}{\Gamma} \exp\left(-.69\left(\frac{\tilde{\nu}-\tilde{\nu}_K}{\Gamma}\right)^2\right)$$  \hspace{1cm} (6)

where $\tilde{\nu}$ and $\Gamma$ are in $\text{kk}$. PROGRAM QMCD, used for matrix method calculations, determines all $R_K$ and $D_K$ then allows for calculations of spectra via eqs. (5) and (6) using a number of different halfwidths in order to search for best agreement with other theories. In any given calculation, however, the same $\Gamma$ is currently used for all polymer bands.

By ignoring doubly excited states, a theoretical problem is created as to whether $\tilde{\nu}_K$ outside the summation or $\tilde{\nu}_i$ inside the summation should be used in eq. (4). Either can be justified, depending on where in the derivation of the theory the following substitution is made.

$$p = \frac{-2\pi mc}{e} \tilde{\nu}$$  \hspace{1cm} (7)

The calculations were done both ways, and the difference, particularly for large polymers, is often extreme. Figure (1) shows calculations of 20 base oligo adenylic acid in RNA geometry using both methods. Only the use of $\tilde{\nu}_K$ gives good agreement with measured
Figure 1. Calculated CD spectrum of oligo A$_{20}$ using the matrix method ($\epsilon_{\text{die1}} = 2$) and, in eq. (4):

$\nu_{K}$ ---; $\nu_{i}$ ---. Both calculations were done with halfwidths of 1.5 kK. Note division by 10 in $\bar{\nu}_{i}$ calculation.
spectra and with calculations done by other theories. It appears that if \( \tilde{\nu}_i \) is used rather than \( \tilde{\nu}_K \), \( C_{iK} \) contributions from distant transitions, e.g. contributions from the 200 nm transitions to polymer \( R_K \) bands in the 260 nm region, become overemphasized because of multiplication by larger values of \( \tilde{\nu}_i \). Bayley, et al.\(^{13}\) point out that this procedure strictly obeys sum rules but is origin dependent, while the opposite is true for use of \( \tilde{\nu}_K \). The choice of \( \tilde{\nu}_K \) versus \( \tilde{\nu}_i \) is inconsequential when only the coupling of closely spaced oscillators is considered, in which case differences in \( \tilde{\nu}_i \) will be of little significance. In the calculations using \( \tilde{\nu}_K \), the \( R_K \) sum nearly to zero anyway, so violation of sum rules appears to be a much less significant problem than origin dependence.

IV. Time Dependent Hartree Theory

A more complete quantum mechanical approach to CD is time dependent Hartree theory.\(^{15,16}\) The Schrödinger equation is constructed from effective base Hamiltonians and is solved reiteratively until self-consistency is reached. In addition to being all order in \( V_{ij} \), frequency dependent monomer susceptibilities can be used to directly calculate the polymer CD as a function of frequency, eliminating the problem of having to assume polymer band shapes. Polymer band rotational strengths directly comparable to the \( R_K \) of
the matrix method or first order perturbation theory cannot in fact be obtained through time dependent Hartree theory, since the contribution of each monomer band to a given polymer band changes with frequency. Nucleic acid time dependent Hartree calculations have so far been limited to an electron in a box model for ApA and oligo $A_N^{17}$ and to sequence dependence studies of hypochromism in double stranded DNA.$^{18-20}$

V. Polarizability Theory

A complete discussion of polarizability theory can be found in two papers by Howard DeVoe, refs. (21) and (22). Rather than using quantum mechanical perturbation techniques, DeVoe returns to a classical coupled oscillator approach. Each monomer transition $i$ is represented by an electric oscillator and a magnetic oscillator which have induced dipoles of the form

$$
\mu_i = \alpha_i (E_i' \cdot e_i) e_i - \left( \frac{\beta_i}{c} \right) (H_i' \cdot e_i') e_i
$$

$$
\mathbf{m}_i = \left( \frac{\beta_i}{c} \right) (E_i' \cdot \mathbf{e}_i) \mathbf{e}_i'
$$

where $E_i'$ is the local electric field at the monomer on which transition $i$ occurs and $e_i', e_i$ are the unit direction vectors for the electric and magnetic dipoles, respectively. $\alpha_i$ is the scalar, frequency dependent monomer polarizability resulting from transition $i$ and is complex in order to account for both absorption and
refraction. \( \beta_i \) is proportional to \( \alpha_i \) by a factor which represents the strength of the magnetic dipole component of transition \( i \) compared to the strength of its electric dipole. In all of the calculations presented here, no transitions with substantial magnetic moments, e.g. \( n - \pi^* \), are considered. Hence, eq. (9) and the second half of eq. (8) can be ignored for purposes of this discussion.

If \( E_0 \) is the electric field at the location of oscillator \( i \) in the absence of all other oscillators, then the local field \( E'_i \) at oscillator \( i \) when the oscillator is placed in an aggregate of oscillators is given by

\[
E'_i = E_0 - \sum_j T_{ij} \cdot \mathbf{e}_j
\]  

where the summation is over all oscillators. \( T_{ij} \) is the point dipole interaction tensor

\[
T_{ij} = |R_{ij}|^{-3}(1 - 3e_{ij}e_{ij})
\]

\( T_{ij} \) is set equal to zero if oscillators \( i \) and \( j \) are on the same monomer. \( R_{ij} = R_j - R_i \) and \( e_{ij} = R_{ij}/|R_{ij}| \).

If there are a total of \( N \) oscillators in the aggregate, there will be \( N \) simultaneous equations

\[
\mu_i \cdot \mathbf{e}_i = \alpha_i \left\{ (E_0 \cdot \mathbf{e}_i) - \sum_j G_{ij} \cdot \mathbf{e}_j \right\}
\]

where \( G_{ij} = e_i \cdot T_{ij} \cdot e_j \) and \( R_i \) is the position vector of oscillator \( i \). \( G_{ij} \) is related to the quantum mechanical
perturbation energy by $G_{ij} = \frac{V_{ij}}{\mu_i \mu_j}$. The solution to this set of equations is a set of $N$ new equations

$$
\mu_i \cdot e_i = \sum_j A_{ij} E_o(R_j) \cdot e_j
$$

(13)

where

$$
A_{ij}(\nu) = \left[ \frac{\delta_{ij}}{a_i(\nu)} + G_{ij} \right]^{-1}
$$

(14)

Each $a_i$ and hence $A_{ij}$ varies with frequency.

The derivation proceeds using Maxwell's equations and the relationship of the complex index of refraction to absorption and CD to give the final results

$$
\epsilon(\nu) = -\left[ \frac{1}{9\eta_s^2} (\eta_s^2 + 2)^2 \right] \left( \frac{8\pi^2 \nu N}{6909} \right)^2 \sum_{i,j} \text{Im} A_{ij} e_i \cdot e_j
$$

(15)

$$
\epsilon_L - \epsilon_R(\nu) = \frac{1}{3300} \left[ \frac{1}{3} (\eta_s^2 + 2) \right] \sum_{i,j} C_{ij} \text{Im} A_{ij}
$$

(16)

where

$$
C_{ij}(\nu) = 24\pi^2 \nu^2 N \left[ (e_i \times e_j) \cdot (R_j - R_i) \right]
$$

(17)

and $\eta_s$ is the frequency dependent index of refraction of the solvent. The bracketed terms in eqs. (15) and (16) have been dropped in all calculations presented here. Although DeVoe's coupled oscillator method is derived by a classical approach, it has been shown to be equivalent to time dependent Hartree theory within the approximations common to all perturbation treatments of polymer optical properties. The overall calculation of CD versus frequency as a function of monomer band shapes is the same, but the formalism is
more straightforward and in general easier to use.

The coupled oscillator equations are strictly correct only if the oscillators can be consistently treated as point dipoles. This is definitely not the case for the calculation of $G_{ij}$ between oscillators on near neighbor bases, since the bases are close together compared to the size of the transition dipoles. In practice, however, any method can be used for calculating $G_{ij}$, since this term is computationally independent of all other terms. In these calculations, the point monopole approximation has been used, giving

$$G_{ij} = \frac{(-4.803)}{\varepsilon_{\text{die}}^2} \sum_{s,t} \frac{q_s^i q_t^j}{|\mathbf{r}_s - \mathbf{r}_t|} / |\mathbf{u}_i| |\mathbf{u}_j|$$  \hspace{1cm} (18)

$G_{ij}$ is in $\text{Å}^{-3}$, $\mathbf{r}$ is in $\text{Å}$, $|\mathbf{u}|$ is in Debye, and $q$ is in esu $\times 10^{10}$. The indices $s,t$ represent all point monopole positions on the bases in which oscillators $i$ and $j$, respectively, occur.

So far, polarizability theory has been used for polynucleotides only to calculate absorption for poly dA:dT and a random sequence of double stranded DNA. Agreement was rather poor, but since a dipole approximation was used, better agreement would have been surprising. Polarizability theory was quite successful, however, in predicting the CD of stereoregular copolymers with aromatic side chains.
VI. The Two-state Dimer: A Comparison of Theories

Some insight is gained by looking at a hypothetical dimer with only one oscillator per base. The equations for each theory can be solved analytically.

When the oscillators are identical, the matrix method and first order perturbation theory results are the same:

\[
R_+ = \frac{\pi \nu_0}{2} R_{12} \cdot \frac{\nu_1 \times \nu_2}{R}
\]

\[
D_+ = |\nu|^2 + \nu_1 \cdot \nu_2
\]

\[
\nu_+ = \nu_0 + \nu_{12}
\]

where \(\nu_+\) has been approximated by \(\nu_0\) in eq. (19). The rotational strengths \(R_+\) and oscillator strengths \(D_+\) are zero order in \(\nu_{12}\), and the splitting is first order. Polarizability theory gives

\[
\epsilon_L - \epsilon_R(\nu) = \frac{3 \times 6909}{3300} \nu \epsilon^0(\nu) \times
\frac{2 G_{12} R R_{12} \cdot \epsilon_1 \times \epsilon_2}{\left[1 - G_{12}^2 (R^2 - I^2)\right]^2 + \left[G_{12}^2 (2IR)\right]^2}
\]

\[
\epsilon(\nu) = \epsilon^0(\nu) \frac{1 - 2G_{12} R \epsilon_1 \cdot \epsilon_2 + G_{12}^2 (I^2 + R^2)}{\left[1 - G_{12}^2 (R^2 - I^2)\right]^2 + \left[G_{12}^2 (2IR)\right]^2}
\]

where \(\epsilon^0(\nu)\) is the monomer absorption band, \(R = \text{Re} \alpha(\nu)\), and \(I = \text{Im} \alpha(\nu)\). Comparison can be made only after band shapes are chosen and optical properties versus frequency calculated from eqs. (19) to (21). For CD, \(R_+\) and \(R_-\) can be assigned band shapes \((\nu_+ f(\nu - \nu_0))\),
respectively, such that $\varepsilon_L - \varepsilon_R$ depends on $V_{12}$ to first order.

$$\varepsilon_L - \varepsilon_R(\bar{\nu}) = f(\bar{\nu} - \bar{\nu}_0) \frac{\pi}{2} \bar{\nu}_0 V_{12} R_{12} \cdot \frac{\mu_1 \times \mu_2}{|\mu_1|^2 |\mu_2|^2}$$

If a further assumption is made that

$$f(\bar{\nu} - \bar{\nu}_0) = \left( \frac{25.1}{\pi \bar{\nu}_0} \right) \frac{\bar{\nu} \varepsilon^0(\bar{\nu}) R}{|\mu_1|^2 |\mu_2|^2}$$

polarizability theory CD equations become equivalent to those of the simpler methods modified only by the denominator in eq. (22). The denominator is symmetric around $\bar{\nu}_0$ and serves to narrow the CD bands, increasing the magnitude near $\bar{\nu}_0$ and decreasing it further away in either direction. The higher order terms in eqs. (22) and (23) are a result of including monomer band shapes in the equations, in effect allowing the vibronic components of the monomer bands to couple independently. DeVoe has shown polarizability theory, truncated to first order, to be analogous to weak coupling perturbation theory.21

When the two oscillators are not degenerate, the matrix method results have a higher order term not present in first order perturbation theory predictions. Letting $\Delta = \bar{\nu}_1^0 - \bar{\nu}_2^0$, the matrix method gives

$$\begin{align*}
\bar{\nu}_+ &= \bar{\nu}_1 - \frac{1}{2} \Delta + \frac{1}{2} \left( \Delta^2 + 4V_{12}^2 \right)^{1/2} \\
\bar{\nu}_- &= \bar{\nu}_2 + \frac{1}{2} \Delta - \frac{1}{2} \left( \Delta^2 + 4V_{12}^2 \right)^{1/2}
\end{align*}$$

(26)
\[
R_+ = -\frac{\pi V_{12}}{(\Delta^2 + 4V_{12})^{1/2}} (R_{12} \cdot \frac{\mu_1 \times \mu_2}{2}) \bar{\nu}^+ \\
R_- = +\frac{\pi V_{12}}{(\Delta^2 + 4V_{12})^{1/2}} (R_{12} \cdot \frac{\mu_1 \times \mu_2}{2}) \bar{\nu}^-
\]

if \(\bar{\nu}_k\) is used in equation (4), or if \(\bar{\nu}_1\) is used,

\[
R^+ = -\frac{\pi V_{12}}{(\Delta^2 + 4V_{12})^{1/2}} \frac{\mu_1 \times \mu_2}{2} (R_2 \bar{\nu}_2^0 - R_1 \bar{\nu}_1^0) \quad (28)
\]

however, Johnson and Tinoco\(^{25}\) have pointed out that if \(\bar{\nu}_1^0\) is close to \(\bar{\nu}_2^0\), so that \((\bar{\nu}_1^0 + \bar{\nu}_2^0)/2\) can be substituted for \(\bar{\nu}_+\), and if the same method used to obtain eq. (24) is applied to eqs. (26) and (27), higher order terms in the matrix method results disappear. The results become identical to first order perturbation theory and are only first order in \(V_{ij}\).

Polarizability theory gives, for non-identical oscillators:

\[
\varepsilon(\bar{\nu}) = \frac{1}{2} \varepsilon_1^0(\bar{\nu}) \times \left[ 1 - 2G_{12} R_2 \frac{\varepsilon_1 \cdot \varepsilon_2}{2} + G_{12}^2 (I_1 I_2 + R_2^2) \right] \frac{1}{\text{Denominator}} \\
+ \frac{1}{2} \varepsilon_2^0(\bar{\nu}) \times \left[ 1 - 2G_{12} R_1 \frac{\varepsilon_1 \cdot \varepsilon_2}{2} + G_{12}^2 (I_1 I_2 + R_1^2) \right] \frac{1}{\text{Denominator}}
\]

(29)

\[
\varepsilon_L - \varepsilon_R(\bar{\nu}) = 12.55\bar{\nu} \left[ \frac{1}{2} \varepsilon_1^0(\bar{\nu}) R_2 + \frac{1}{2} \varepsilon_2^0(\bar{\nu}) R_1 \right] \\
\times \left[ \frac{G_{12} R_{12} \cdot \mathcal{S}_1 \times \mathcal{S}_2}{\text{Denominator}} \right]
\]

(30)

where

\[
\text{Denominator} = \left[ 1 - G_{12}^2 (R_1 R_2 - I_1 I_2)^2 \right]^2 + \left[ G_{12}^2 (R_1 I_2 + R_2 I_1)^2 \right]^2
\]
If the oscillators are widely separated in frequency, Im \( \alpha \) for oscillator 2 will be zero and Re \( \alpha \) for oscillator 2 only slightly sloped in the frequency range for which oscillator 1 has non-zero Im \( \alpha \). Both the CD and absorption bands arising from oscillator 1 will be scaled versions of \( \varepsilon^0(\nu) \) for oscillator 1. In other words, the distant oscillator can significantly affect the magnitude but not the position or shape of a band arising from oscillator 1.

VII. Computational Considerations

The matrix method program, QMCD, and the polarizability theory program, ROTOPM, are described and listed in the appendix. Equations (22) and (23) were not used directly in ROTOPM for calculating the isotropic properties. Rather, \( A_{ij} \) was diagonalized to obtain components along its principal axes. The CD for light incident parallel and perpendicular to the helix axis could then be calculated as well as the isotropic CD.

All computations were carried out on the Lawrence Berkeley Laboratory CDC 7600 computer. Computer size limits the polymer size that can be handled only with respect to the total number of oscillators contained in the polymer. The complex matrix \( A_{ij}(\nu) \) is of dimension \( NxNx2 \) where \( N \) is the total number of oscillators. Both ROTOPM and QMCD are currently dimensioned
to accept a maximum of 140 oscillators. Therefore, a polymer of 20 bases (10 base pairs), each with 7 oscillators, or 70 bases, each with 2 oscillators, can be calculated. Currently, ROTOPM requires a field length of 157,400 words (octal) to load and run whereas QMCD requires 143,200 if individual gaussians are not printed out. Current 7600 limit is 170,000.

Typical time requirements for a 140 oscillator calculation are approximately 200 CPU seconds (700 CUS) for 26 points on ROTOPM and 100 CPU seconds (340 CUS) for 76 points on QMCD. For ROTOPM, the number of frequency points calculated is an important determining factor in the cost of large polymer calculations, since a new $A_{ij}$ matrix must be set up and inverted at each frequency.

VIII. A Note on Intrinsic Monomer CD

None of the methods for calculating polymer CD presented here take into account the intrinsic CD of the mononucleotides. To compensate for this, calculated CD spectra could be compared with measured spectra from which monomer CD has been subtracted. This has not been done. In many cases, particularly when the polymer CD is large, the correction is small anyway. Subtracting the CD of the adenine ribonucleoside from ApA, measured under identical conditions, does not change any peak or trough by more than $\varepsilon_L - \varepsilon_R = 1$. 
The deoxy correction is even smaller. However, a totally different monomer correction might be required if the monomer conformation changes or becomes more rigid when part of a polymer, since intrinsic CD is expected to be sensitive to base-sugar-phosphate conformation. Such a change in monomer CD is difficult to predict, so presuming the intrinsic CD will remain fairly small we will disregard it.
CHAPTER 1 REFERENCES


Chapter 2

MONOMER INPUT PARAMETERS

I. Number and Location of Monomer Transitions

Probably the most difficult problem and largest source of error in calculating polynucleotide optical properties is determining the number and location of monomer absorption bands. For absorption above 220 nm, the choice was influenced by dichroic ratio measurements on crystals and oriented films\(^1\)\(^{-8}\), CD\(^9\) and MCD\(^10\) measurements on monomers in aqueous solution, and predictions made by CNDO-CI molecular orbital calculations.\(^11\) All resolutions were done on absorption spectra of mononucleotides in neutral solution.\(^12\) The uridylic acid spectrum was resolved into only one band, at 262 nm, an assumption supported by the CNDO calculations and all dichroic ratio studies except recent polarized single-crystal reflection experiments on 6-azauracil.\(^13\) Absorption spectra of the other three mononucleotides were each resolved into two bands above 220 nm. The guanylic acid absorption spectrum is clearly composed of two bands, but the resolution was somewhat arbitrary since a number of resolutions resulting in different ratios for the oscillator strengths of the two bands could be made. Adenylic acid has been traditionally assigned a large band at 260 nm and a much smaller band at 240 nm, although evidence exists
that the smaller band may be located elsewhere, perhaps even on the red side of 260 nm. (See ref. (8) for review of evidence.) The cytidylic acid spectrum was the most difficult to resolve. One band is clearly located at 270 nm, but because of a shoulder in the absorption spectrum at 215 nm, fitting two bands between 210 and 260 nm gave the best resolution. In the absence of any strong evidence for more than one band in this region, however, a single band was resolved at 232 nm.

Crystal and film studies of the bases have not provided resolutions of bands below 200 nm. Absorption bands were assigned to the 180 to 210 region on the basis of the CNDO calculations, trying to account for the absorption with as few oscillators as possible. For adenine, guanine, and uracil, the CNDO calculations show the polarizability in this region to be about half parallel and half perpendicular to the longest wavelength transition, so two mutually perpendicular bands were resolved here for each base. For cytosine, only one strong band is predicted in this region, polarized at about $45^\circ$ to the first band. The calculated CD near 260 nm is expected to be affected somewhat differently by these bands located relatively close to 260 than it would be by including them in the background polarizability. They will provide a small overlap in Im $\alpha$ and a more sloped Re $\alpha$ tail above 230 nm.
Resolution of the measurable absorption for all four bases are shown in Figures (2a,b,c,d).

Finally, three mutually perpendicular background polarizabilities, one out of plane, were placed arbitrarily at 119 nm as a means of taking into consideration the vacuum UV absorption of each mononucleotide. Table (1) gives a summary of specifications for all bands.

Except for one of the background oscillators, no out of plane transitions, e.g. n - π*, were explicitly taken into account. The effect of these transitions on the observed optical properties above 220 nm of polymers has never been satisfactorily determined. An important goal of these calculations is to see how completely the CD can be explained solely by the interactions of π - π* transitions.

The sugar-phosphate backbone was also ignored. Calculations by Moore and Wagner support the assumption that the calculated CD above 220 nm will depend very little on backbone transitions since they are distant in frequency as well as space.

II. Polarizabilities

For the absorption bands above 180 nm, polarizabilities were computed from the extinction coefficient as a function of wavelength by a Kronig-Kramers transform according to the equations
Figure 2. Resolutions of mononucleotide absorption spectra. (All spectra from Voet et al.\textsuperscript{12}): (a) adenosine-3'-(2')-phosphoric acid, pH 7.9; (b) uridine-3'-(2')-phosphoric acid, pH 7.6; (c) guanosine-3'-(2')-phosphoric acid, pH 7.7; (d) cytidine-3'-(2')-phosphoric acid, pH 7.9. Numbering for the purines is shown in Figure 22.
(a) Adenylic Acid
(c) Guanylic Acid
Table 1

<table>
<thead>
<tr>
<th>λ_{max} (nm)</th>
<th>\epsilon_{max} \times 10^{-3}</th>
<th>\mid \mu \mid \text{in Debye}</th>
<th>\theta</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 260</td>
<td>15.0</td>
<td>3.78</td>
<td>177°</td>
</tr>
<tr>
<td>240</td>
<td>3.1</td>
<td>1.57</td>
<td>87°</td>
</tr>
<tr>
<td>207</td>
<td>20.6</td>
<td>4.42</td>
<td>87°</td>
</tr>
<tr>
<td>187.5</td>
<td>18.1</td>
<td>4.17</td>
<td>-3°</td>
</tr>
<tr>
<td>119</td>
<td>10.79</td>
<td>87°</td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>9.26</td>
<td>177°</td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>7.98</td>
<td>\perp</td>
<td></td>
</tr>
<tr>
<td>U 262</td>
<td>10.0</td>
<td>3.20</td>
<td>171°</td>
</tr>
<tr>
<td>206</td>
<td>8.85</td>
<td>2.40</td>
<td>127°</td>
</tr>
<tr>
<td>175</td>
<td>16.0</td>
<td>4.37</td>
<td>-33°</td>
</tr>
<tr>
<td>119</td>
<td>6.92</td>
<td>-9°</td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>6.70</td>
<td>81°</td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>6.66</td>
<td>\perp</td>
<td></td>
</tr>
<tr>
<td>G 281</td>
<td>6.5</td>
<td>2.22</td>
<td>176°</td>
</tr>
<tr>
<td>252</td>
<td>13.6</td>
<td>4.19</td>
<td>-75°</td>
</tr>
<tr>
<td>202.5</td>
<td>16.0</td>
<td>4.22</td>
<td>-74°</td>
</tr>
<tr>
<td>186</td>
<td>21.4</td>
<td>4.64</td>
<td>-6°</td>
</tr>
<tr>
<td>119</td>
<td>10.31</td>
<td>86°</td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>8.85</td>
<td>176°</td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>7.98</td>
<td>\perp</td>
<td></td>
</tr>
<tr>
<td>C 270</td>
<td>8.70</td>
<td>3.00</td>
<td>12°</td>
</tr>
<tr>
<td>232.5</td>
<td>7.95</td>
<td>3.43</td>
<td>0°</td>
</tr>
<tr>
<td>198</td>
<td>21.9</td>
<td>4.82</td>
<td>-90°</td>
</tr>
<tr>
<td>119</td>
<td>6.54</td>
<td>0°</td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>6.34</td>
<td>90°</td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>6.66</td>
<td>\perp</td>
<td></td>
</tr>
</tbody>
</table>
The Kronig-Kramers program, CDORD, is listed in the Appendix. The original program by M. Itzkowitz, which does the transform numerically by increments of wavelength, was patched to accept data as $\varepsilon$ versus $\lambda$, to perform the transform, and to interpolate the result to give output as both $\alpha$ versus $\lambda$ and $\alpha$ versus $\tilde{\nu}$. The latter is required for PROGRAM ROTOPM.

For the background oscillators, the contribution of all other bands to the polarizability at 589 nm (Na D line), i.e. the sum of Re $\alpha_i(589)$ from eq. (32) for each of these bands, was subtracted from the total measured polarizability at 589 for the model compounds quinoline and pyridine. The remaining polarizability, split by the measured anisotropy ratios of the model compounds, was used in Lorentzian equations to determine each of the background bands.

\[
\text{Im } \alpha_i(\tilde{\nu}) = \left| \mu_i \right|^2 \frac{2}{hc} \frac{\tilde{\nu} \Gamma_i}{(\tilde{\nu}_i^2 - \tilde{\nu}^2)^2 + \tilde{\nu}_i^2 \Gamma_i^2} \quad (33)
\]

\[
\text{Re } \alpha_i(\tilde{\nu}) = \left| \mu_i \right|^2 \frac{2}{hc} \frac{\tilde{\nu} (\tilde{\nu}_i^2 - \tilde{\nu}^2)}{(\tilde{\nu}_i^2 - \tilde{\nu}^2)^2 + \tilde{\nu}_i^2 \Gamma_i^2} \quad (34)
\]

The actual shape and strength of each band depends on the choice of frequency, maximum and half width, $\Gamma_i$, but the parameters are unimportant except as they determine the tail of the polarizability above 200 nm.
This tail is only slightly sloped and is not very sensitive to changes in the variable parameters.

Several tedious steps would be avoided if eqs. (33) and (34) could be used for determining the polarizability of all absorption bands. Unfortunately, the near UV bands are very nearly gaussian and are fit by the Lorentzian equations extremely poorly.

For matrix method calculations, $|\mu_i|$ and $|\mu_i|$ are needed rather than $a_i(\vec{v})$. A simple integration of $\varepsilon_i(\lambda)$ for each band, using PROGRAM FINALE, provided $|\mu_i|$. $\vec{v}_i$ was chosen to be the frequency of maximum absorption for each band.

III. Transition Dipole Moments

Directions for the $\varepsilon_i$ were taken when available from measurements on the individual bases. For the transitions in the 260 nm region, values of $\theta$ (defined in Figures 2a and 2b) were taken from crystal absorption studies of adenine, from crystal absorption and polarized reflection spectra of uracil and thymine, from crystal polarized reflection spectra and single crystal specular reflectance measurements of guanine, and from crystal absorption and specular reflectance studies of cytosine. Specific data is available for transitions in the 200 nm region only for guanine, uracil and thymine. Stretched film studies show the rise in absorption below 220 nm for all four bases to
be oriented approximately perpendicular to the longest wavelength transition.

For the purines, the larger of the two in-plane background oscillators was assigned the same value for $\theta$ as the longest wavelength transition, since it is expected that the long axis of an aromatic molecule will be more polarizable than the short axis. Correspondingly, the other in-plane oscillator was oriented in a perpendicular direction. For the pyrimidines, the larger of these two oscillators was arbitrarily assigned the same value for $\theta$ as the longest wavelength transition and a value for $\theta$ perpendicular to the first for the other in-plane oscillator. Values of $\theta$ that were used in the calculations are listed in Table (1) for all oscillators.

Each oscillator was placed at a position $R_i$ dependent upon the distribution of its transition monopoles by the equation

$$R_i = \sum_{\text{atoms } s} |q_s| \frac{\mathbf{r}_s}{\sum_s |q_s|}$$

where the $\mathbf{r}_s$ are the atom position vectors of the base on which oscillator $i$ is located. When CNDO calculated monopoles (see below) were used, this position often was far removed from either the geometric or weighted center of the base.
IV. Monopoles

The transition monopoles used to calculate $G_{ij}$ by eq. (18) are defined such that

$$ \mu_i = -4.803 \sum_s q_s^i r_s $$  \hspace{1cm} (36)

where $\mu_i$ is in Debye, $r$ is in Å, and $q$ is in esu x $10^{10}$. It should be noted that the direction of $\mu_i$ for any oscillator can be changed by $180^\circ$ as long as the signs of its monopoles are reversed. However, the monopole signs of all oscillators must be consistent with the oscillator directions by the same definition, in this case eq. (36).

Transition monopoles for the first two oscillators of all bases were taken from CNDO-CI molecular orbital calculations. The $\theta$ values for the dipoles calculated from these monopoles by eq. (36) do not agree directly with the measured $\theta$ used in the calculations. (Compare values of $\theta$ listed in Tables 1 and 2.) Calculations were normally done, nevertheless, using the original CNDO monopoles. To keep $\theta$ as consistent with the monopoles as possible, $\theta$ or $\theta + 180^\circ$ was chosen, depending on which was closer to the monopole derived $\mu_i$. These monopoles, listed in Table (2), are scaled to give the measured $|\mu_i|$ for the corresponding resolved absorption band. For polarizability theory calculations, any scaling factor can be used that is consistent with the value of $|\mu_i|$ used in
Table 2

|   | \( \theta \) | \( |\Delta \mathbf{S}_{ij}(\text{Debye})| \) | N1   | C2   | N3   | C4   | C5   | C6   | N7   | C8   |
|---|------------|---------------------------------|------|------|------|------|------|------|------|------|
| I | -133.76    | 3.78                            | -0.088700 | .117700 | -0.085900 | -0.028600 | -0.018700 | .102500 | .256700 | -0.402100 |
| II| 132.74     | 1.57                            | -0.007400 | .118600 | .086100 | -0.287300 | .210400 | .047200 | -1.01300 | .087100 |
| (I)|            |                                 | .069600 | .077500 |
| (II)|           |                                 | .065600 | .018200 |
| G | -136.52    | 2.22                            | -0.025200 | .012000 | -0.018100 | -0.048800 | .056900 | .015100 | .163800 | -0.272100 |
| II| -57.28     | 4.19                            | .011200 | .190100 | -1.50100 | .388900 | -0.328600 | -0.032900 | .140800 | -0.197000 |
| (I)|            |                                 | .057300 | .008500 | .050600 |
| (II)|           |                                 | .051900 | .075900 | -0.046400 |
| U | 176.16     | 3.20                            | -0.073100 | .012600 | -0.012200 | -0.029400 | .313200 | -2.66800 | -0.018100 | .073800 |
| II| 140.40     | 2.40                            | .036800 | .074500 | -0.094700 | -1.125800 | -0.093000 | .086800 | -1.38300 | .253700 |
| C | 16.11      | 3.00                            | .161500 | -0.020000 | -0.163000 | .061000 | -0.310000 | .265000 | -0.001900 | .012400 |
| II| -25.40     | 3.43                            | .161600 | -0.046300 | .301000 | -1.201000 | .000600 | -0.053200 | -0.234000 | -0.220200 |

**NOTE:** \( \theta \) represents the transition angle as calculated by eq. (36) from these \( \mathbf{S}_{ij} \).
calculating $G_{ij}$ by eq. (18).

CNDO monopoles were not used for the remaining oscillators, since direct correspondence between oscillator and real transitions is no longer certain. For in-plane oscillators a method was devised by which delocalized monopoles were placed on six atoms of each base such that they gave the desired $\theta$. PROGRAM BASES, listed in the Appendix, was designed for this purpose as well as for calculating $R_i$ from the atom coordinates and monopoles and for calculating the spherical coordinates of $e_i$ from the atom coordinates and $\theta$. The out-of-plane background oscillator was assigned monopoles $0.75 \, \text{Å}$ above and below each atom of the base, a distance calculated from carbon and oxygen Slater orbitals. Each out-of-plane monopole was assigned a value $q_s^i = \pm |\mu_i|/(2 \times \text{number of atoms})$, with the signs all positive or all negative on either side of the base in accordance with eq. (36) and the chosen $\theta$ for $\mu_i$. Monopoles for adenine, calculated by PROGRAM BASES using RNA coordinates, are listed in Table (3) as examples.

V. Geometry

Coordinates for the bases were taken from X-ray diffraction studies of DNA for B-DNA\textsuperscript{16} and C-DNA\textsuperscript{17} geometries and of poly rA:rU for RNA geometry.\textsuperscript{18} In single strand calculations, the bases are assumed to
Table 3

BASES Calculated Monopoles for Adenine in RNA Geometry

| Osc | $\theta$ | $|\mu_i|$ | N1 | C2 | N3 | C4 | C5 | C6 | N7 |
|-----|---------|--------|-----|----|----|----|----|----|----|
| A   | 3 87°   | 4.47   | 0.000000 | 0.000000 | -161770 | 0.000320 | -0.56210 | -1.82010 | 1.179690 |
|     | 4 -3°   | 4.17   | 0.000000 | 0.000000 | -1.29340 | 0.225750 | -2.15600 | -1.05470 | -1.120290 |
|     | 5 87°   | 10.79  | 0.000000 | 0.000000 | -3.39420 | 0.000700 | -1.37710 | -4.29430 | 0.438540 |
|     | 6 177°  | 0.26   | 0.000000 | 0.000000 | -2.87710 | -5.01310 | 0.478740 | 0.234208 | 0.254100 |


\[
\begin{align*}
\text{Table 3 (cont.)} \\
\text{BASES Calculated Monopoles for Adenine in RNA Geometry} \\
\end{align*}
\]

\[\begin{array}{c|c|cc}
(3) & C8 & N9 & (NH_{2})_6 \\
& 0.000000 & 0.217200 & 0.000000 \\
(4) & 0.000000 & 0.984200 & 0.000000 \\
(5) & 0.000000 & 0.521300 & 0.000000 \\
(6) & 0.000000 & 0.191550 & 0.000000 \\
\end{array}\]

Out-of-plane $\theta = 7.8^\circ$ all monopole positions above plane of base
$\theta = 110.8^\circ$ all monopole positions below plane of base
Table 4

GEOMETRY SPECIFICATIONS

<table>
<thead>
<tr>
<th></th>
<th>Axial Increment (Å)(^a)</th>
<th>(\theta_b) (deg)</th>
<th>Residues per Turn</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-RNA</td>
<td>2.81</td>
<td>32.7</td>
<td>11</td>
</tr>
<tr>
<td>B-DNA</td>
<td>3.38</td>
<td>36.0</td>
<td>10</td>
</tr>
<tr>
<td>C-DNA</td>
<td>3.32</td>
<td>38.6</td>
<td>9.33</td>
</tr>
<tr>
<td>Arnott et al(^c)</td>
<td>3.03</td>
<td>45</td>
<td>8</td>
</tr>
<tr>
<td>poly AU:AU and poly GC:GC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arnott et al(^c)</td>
<td>3.26</td>
<td>30.0</td>
<td>12</td>
</tr>
<tr>
<td>poly dT:dA:dT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arnott et al(^d)</td>
<td>3.41</td>
<td>31.3</td>
<td>11.15</td>
</tr>
<tr>
<td>4 strand poly I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rich(^d)</td>
<td>3.40</td>
<td>41.6</td>
<td>8.68</td>
</tr>
<tr>
<td>3 strand poly I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zimmerman et al(^d)</td>
<td>3.36</td>
<td>31.2</td>
<td>11.21</td>
</tr>
<tr>
<td>4 strand poly I</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(a\). Distance along helix axis between adjacent bases
\(b\). Angle between adjacent bases
\(c\). See Chapter 4
\(d\). See Chapter 6
be stacked upon one another as in one of the two strands of the double stranded polymers. References will be stated in the text when non-standard geometries, e.g. 4-stranded poly I, are used. Descriptive parameters for the various geometries used are listed in Table (4).
CHAPTER 2 REFERENCES

   Phys. 39, 255-266.
   Phys. 40, 2071-2075.
   53, 2164-2172.
   J. Amer. Chem. Soc. 93, 6679-6680.
   Chem. Soc. 92, 3593-3599.
   Soc. 93, 2054-2056.
   Chem. Soc. 93, 6443-6448.
10. Voelter, W., Records, R., Bunnenberg, E., and
    Soc. 95, 2803-2813.
12. Voet, D., Gratzer, W.B., Cox, R.A. and Doty, P.
    (1963) Biopolymers 1, 193-208.
    B.G. (1974) J. Amer. Chem. Soc. 96, 1597-
    1598.
    13, 977-986.
    of Pure and Applied Chemistry 5, 261-318.

Chapter 3

ApA AND OLIGO A_N CALCULATIONS

I. Objective

Calculations will first be presented for adenylyl-(3'→5')-adenosine (ApA) and oligo adenylic acid (oligo A_N) of varying chain length in both RNA and B-DNA geometries. ApA, both ribo and deoxy forms, and poly rA (but perhaps not poly dA) are all thought to be stacked in conformations very similar to one of the two strands of a double stranded polymer. (See Brahms and Brahms for review of evidence.) While reasonably accurate calculations for these polymers in the assumed geometries have already been published, the relative simplicity of calculations involving such uncomplicated sequence and structure provides a good opportunity for assessing the advantages of polarizability theory. A slight modification of the theoretical procedure to include an effective dielectric constant will also be evaluated. Empirical variations in the number and properties of monomer absorption bands will be made in an attempt to determine what types of parameter changes might be expected to correct remaining discrepancies between calculated and measured spectra.

II. ApA Calculations

Figure (3a) shows the CD spectrum for ApA calculated in RNA geometry using polarizability theory and
Figure 3. (a) CD spectrum of ribo ApA: calculated in RNA geometry ———; measured at 26°C in 0.01 M Na–PO$_4$ buffer, 0.1 M NaClO$_4$, pH 7.2 ———; (b) CD spectrum of deoxy ApA: calculated in B-DNA geometry ———; measured at 26°C in 0.01 M Na–PO$_4$ buffer, 0.1 M NaClO$_4$, pH 7.2 ———. Calculations were done with $\epsilon_{diele} = 2$. 
(a) ribo ApA

(b) deoxy ApA

Wavelength (nm)
the monomer properties outlined in Chapter 2. A measured CD of ribo ApA at 26°C (0.01 M Na-PO₄ buffer, 0.1 M NaClO₄ at pH 7.2)⁶ is shown for comparison. At this temperature, where the CD probably represents the average of many structural fluctuations, the calculated CD is in good agreement with that measured, although the calculated magnitude at 270 nm is about two times too large. Measurements done at lower temperature, e.g., at -1.5°C in 0.1 M NaCl, 0.1 M Tris⁷ and at -68°C at high LiCl⁸ concentration indicate that when internal motion of the dimer is hindered, the measured magnitudes of both the 260 region peak and trough approximately coincide with the calculated magnitudes. The -1.5°C measurement, for instance, shows a maximum of approximately 10.8 at 270 nm and a minimum of -9.2 at 250 nm. Similar calculations for ApA in B-DNA geometry are shown in Figure (3b) along with a 26°C measurement of the deoxy ApA CD spectrum.⁹ While exact agreement is not obtained, the minor differences in measured CD shape between ribo and deoxy forms are predicted by the calculated CD.

The calculated absorption spectra for ApA in RNA and B-DNA geometries also agree reasonably well with measurement. There is little discernable difference between the measured absorption spectra of ribo and deoxy ApA, both having a maximum extinction coefficient of roughly 13.8 x 10³ near 257 nm.⁷,10-12
For ribo ApA, the calculated $\lambda_{\text{max}}$ is about 256 nm, only slightly blue shifted from the measured maximum, and for deoxy ApA, the calculated maximum is at 257 nm. The calculated $\varepsilon_{\text{max}}$ is about $13.0 \times 10^3$ for both RNA and B-DNA geometries. This is a bit low, but the exact amount of hypochromism at 260 nm is expected to be difficult to predict since it depends solely on borrowing from the less well understood higher energy monomer absorption bands. Interestingly, the measured absorption spectra of ApA in high LiCl concentration has an $\varepsilon_{\text{max}}$ value of about $13.0 \times 10^3$.

For purposes of comparison and to look more closely at the polymer band structure, the same calculations were done with the matrix method. The calculated $R_X$ pattern for ApA in RNA geometry (Figure (4)) shows that for such a dimer, the CD arises almost totally from a series of equal but opposite in sign $R_X$ pairs. The two dimer transitions in each pair appear to arise mainly from interaction between identical monomer transitions on the two bases, with only small contributions from other transitions. This is borne out by the coefficients of monomer band contribution to the eigenvectors and by the fact that there is little anisotropy within each $R_X$ pair. A direct comparison between measurement and calculation is difficult to make for matrix method results, because the resultant CD shape depends on the choice of band shape arbitrarily as-
Figure 4. Matrix method CD calculations for (a) ApA and (b) oligo A20 in RNA geometry with ε_{diel} = 2. $R^*_K$ patterns are shown by rectangles (positive bands solid, negative bands open) centered at $\tilde{\nu}_K$. $R^*_K$ values have been added when splitting is too small to show. CD spectra were calculated using gaussian band shapes of halfwidth 1.5 kK. $R^*_K$ values are per molecule rather than per nucleotide.
(a) ApA

(b) oligo A_{20}

Wavelength (nm)
signed to each calculated rotational strength. In
the spectrum shown, remarkably good agreement with the
polarizability theory result is obtained by using
gaussians of bandwidth 1.5 kK for each polymer band.
If the bandwidth is decreased, the long wavelength CD,
for instance, will be less broad with a maximum and
minimum closer to 260 nm but with larger magnitude.
Somewhat different spectra would be obtained using
other choices of band shapes.

III. Oligo Aₙ Calculations

Using the same input parameters the polarizability
theory dimer calculations were extended in chain length
up to a maximum of 20 bases. Calculated CD spectra
of oligo Aₙ in RNA geometry for selected values of N
are shown in Figure (5a). Measured CD spectra for
oligo A of N=3 and 6 and of poly A, all in 0.1 M NaCl,
0.01 M Tris at pH 7.4 (D. Gee, unpublished work), are
shown in Figure (5b). (These measured CD spectra were
determined using values for ε max near 257 nm of 12.9
x 10³, 12.0 x 10³, and 10.3 x 10³, respectively.) The
calculated chain length dependence of the 265 nm band
follows measurement remarkably well. Calculated
maximums are 14.8 at 266.5 nm for A₃, 17.2 at 265.5
for A₆, and 18.6 at 264.5 for A₂₀. Comparable measured
spectra show values of 12.4 at 270nm for A₃, 19.5 at
268 for A₆, and 22.6 at 266.5 for poly A. The mea-
Figure 5. (a) Calculated CD spectra of oligo A$_N$ in RNA geometry for $N = 2, 3, 6,$ and $20$ using $\epsilon_{\text{diel}} = 2$.
(b) Measured CD spectra of ribo A$_3$, ribo A$_6$ and poly rA at $0^\circ C$ in 0.1 M NaCl, 0.01 M Tris, pH 7.4
(D. Gee, unpublished work.)
(a) Oligo $A_N$: calculated

(b) measured
sured magnitude of the 265 nm band shows temperature
dependence and some variation with salt conditions
(compare spectra in refs. 11 and 13), so it is diffi-
cult to interpret these results. Nevertheless, it is
tempting to hypothesize from the calculations that
increase in chain length decreases the flexibility of
the polymer toward a more rigid RNA-like structure.
The maximum extinction coefficient for $A_{20}$ is calcu-
lated to be about $11.6 \times 10^3$ at 253 nm. Published
values for $\epsilon_{\text{max}}$ of poly A also vary, ranging from $9$
$\times 10^3$ at 256.5$^{11}$ to $10.5 \times 10^3$ at 257 nm$^{14}$ with shorter
chain length having higher absorption. The calculated
wavelength maximum is reasonable, and the calculated
magnitude is perhaps as good as can be expected con-
sidering the lack of accurate information for far UV
transitions.

Agreement with measured CD at shorter wavelengths
is somewhat less satisfactory, the peak/trough ratio
being correct only for $A_6$ and the positive band at
215 nm having been completely lost for chain lengths
greater than 3. The matrix method gives a CD spectrum
for $A_{20}$ (Figure (4)) very similar to polarizability
theory results, and inspection of the $R_X$ pattern pro-
vides possible insight into the sources of error. Ro-
tational strengths arising from both the 240 and 207 nm
monomer oscillators are very weakly split. The CD
shape and magnitude in these regions is therefore much
more dependent on the relative strengths of the positive and negative $R_X$ bands than in other regions, i.e., borrowing from other monomer bands is much more influential. Large negative rotational strengths from the 207 nm oscillator appear at slightly higher wavelengths than the large positive bands, an inversion of the dimer pattern, so a larger splitting would still give the wrong CD sign near 210 nm. Correction of the calculated CD in this region will apparently depend on changing the interaction of the 207 nm oscillator with its neighbors.

The calculations of oligo $A_N$ in B-DNA geometry are very similar to the RNA results in the 260 region, although the change in shape and magnitude with increase in chain length is much smaller and appears to approach a limit more rapidly. The measured CD for poly dA,$^{13}$ on the other hand, bears little resemblance to the calculated CD. Above 230 nm, the measured CD of poly dA is unlike deoxy ApA, having two very small positive CD bands with maxima at 278 and 265 nm. Considering the reasonable success of the RNA calculations above 230 nm, it seems improbable that poly dA is in any geometry close to standard B-DNA. Below 230 nm, the calculated CD for $A_{20}$ in B-DNA geometry differs radically from the RNA calculation in that the 207 nm monomer oscillator produces a very large positive band ($\varepsilon_L - \varepsilon_R$ about 18) at 213 nm and a much
larger negative band at 202 nm, clearly from exciton-type coupling.

IV. Dependence of Calculations on Effective Dielectric Constant

All of the above calculations were done using an effective dielectric constant of 2. In Figure (6), the calculated CD of ApA and oligo A$_{20}$ in RNA geometry using $\epsilon_{\text{diele}}$ equal to 1, 2, and 4 are compared. Changing $\epsilon_{\text{diele}}$ from 1 to 2 in the ApA CD calculations causes small shifts in extrema and crossover wavelengths and reduces the magnitude more or less by 2. For oligo A, the magnitude is reduced by much less than 2, and the shape changes significantly. Agreement with measurement is improved above 230 nm while the positive band at 215 nm is lost. Introducing $\epsilon_{\text{diele}} = 2$ in CD calculations has proven to give unpredictable results whenever the polymer involves more than two bases. Increasing $\epsilon_{\text{diele}}$ to higher values generally just reduces the CD magnitude, particularly for dimers where the CD becomes nearly exactly proportional to $G_{ij}$. Successive increases in $\epsilon_{\text{diele}}$ change the absorption near 260 nm of both ApA and oligo A by increasing the maximum extinction coefficient and red shifting the wavelength maximum.

Differences between polarizability theory and matrix method calculations are more pronounced when $\epsilon_{\text{diele}} = 1$ is used. In the RNA calculation of oligo A$_{20}$, for instance, the positive peak at 217 nm cal-
Figure 6. Calculated CD spectra of (a) ApA and (b) oligo A20 in RNA geometry using $\epsilon_{\text{diel}} = 1 \ldots$, $\epsilon_{\text{diel}} = 2 \ldots$, and $\epsilon_{\text{diel}} = 4 \ldots$. 
culated by polarizability theory (Figure 6) is totally missing in the matrix method calculation.

V. Variation of Monomer Parameters

To determine whether better agreement with measured spectra could be obtained for both ribo ApA and oligo A_{20}, a number of trial calculations testing variations in parameters were done with polarizability theory.

(A) Number of Oscillators

Figure (7) shows a series of calculations in which one or more of the 7 oscillators per base are omitted. It is apparent that in polarizability theory as well as the matrix method, the CD pattern in a given wavelength region is determined mainly by the interaction of identical oscillators in that region, with the exception of the oligo A_{20} CD arising from the 207 nm oscillator. Polymer bands generally change in magnitude only, not in position or shape, when oscillators at distant wavelengths are eliminated. It should be noted that the direction the magnitude changes, e.g., for the longest wavelength positive CD band, is not necessarily the same for ApA and A_{20} due to the unpredictable influence of non-nearest neighbor interactions. None of the eliminations clearly improve the calculated spectra. Omitting only the out-of-plane background oscillator (not shown) improves the 260
Figure 7. Calculated CD spectra of (a) ApA and (b) oligo A20 in RNA geometry using $\varepsilon_{\text{diel}} = 2$ and all 7 oscillators -----; no background oscillators ---; oscillators at 240 and 260 nm only ······.
peak/trough ratio for $A_{20}$, but has the same effect on $ApA$, further distorting the CD. Eliminating the background altogether in $A_{20}$ corrects the 215 nm CD somewhat, but not enough. The absorption spectra for these calculations show little or no shift in $\lambda_{\text{max}}$, only changes in magnitude. As expected, the hypochromism of the 260 nm region decreases with successive elimination of far UV and background oscillators. CD calculations for B-DNA show an even smaller dependence of the 260 nm bands on other oscillators, but occasionally show larger changes in the 200 region.

(B) Variations in $\theta$

Rather than eliminating oscillators from the calculations, attempts were made to correct the CD spectra by varying $\theta$ for the non-background oscillators. PROGRAM ROTOPR (see Appendix) was used for this purpose. Monopoles were not changed, so $\theta$ for each oscillator was kept within $90^\circ$ of the angle calculated from its monopoles by eq. (36). Results were essentially as expected. Variations of $\theta$ for any one oscillator did little more than change the magnitude of peaks arising from other oscillators. The only exception was that bands resulting from the 187.5 and 207 nm oscillators were somewhat interdependent, in agreement with the fact that these monomer absorption bands are both large and close together. The CD aris-
ing from the oscillator for which $\theta$ was being varied was also sensitive, usually just in magnitude, to the choice of angle since the monopole rather than point dipole approximation is being used. More sensitivity was found in oligo $A_{20}$ calculations than in ApA. Occasionally better peak to trough ratios, for instance, were produced in one wavelength region at the expense of good agreement in another. Overall, no new combination of transition angles was found that provided significantly better calculations for both ApA and oligo A.

A similar test on the effect of background oscillators was achieved by going to extremes in the in-plane anisotropy ratio. Putting all in-plane background strength into one oscillator accentuates CD bands and increases hypochromism of absorption bands coming from oscillators parallel to the chosen $\theta$ of the new background oscillator. The opposite effect is seen on bands from perpendicular oscillators. The most extreme effects were observed assigning all in-plane background to $\theta = 177^\circ$ where, in ApA, the CD in the 220 nm region became totally negative, and in oligo $A_{20}$ the negative band at 250 nm disappeared. No beneficial shifts in CD bands were ever produced.

(C) Adjusting the Monopoles

Unfortunately, while the method of determining
monopoles may be one of the most important determinants of the calculated CD pattern, there is no easy way of systematically varying the monopoles. The practice of using CNDO calculated monopoles which do not give rise, via eq. (36), to the $\theta$ used in the calculations is particularly questionable. An attempt was made to alter the monopoles for the 260 and 240 nm oscillators by calculating monopoles for the difference dipole $\mu = \mu_{\text{meas}} - \mu_{\text{CNDO}}$ and adding them to the CNDO monopoles. Calculations were also done using monopoles for these two oscillators calculated by the same method normally used for the remaining oscillators. In both cases, new peak/trough ratios were observed, but the original CNDO monopoles, uncorrected for $\theta$, gave consistently better agreement with measured spectra. On the other hand, it is quite possible that a different distribution of monopoles for the 207 nm oscillator might give better results for the $A_{20}$ CD spectrum near 220 nm. A trial calculation in which the 260 nm oscillator monopoles were used for the 207 nm oscillator as well gave similar $R_K$ patterns near 260 and 207 nm, but the final CD was much too large at short wavelengths and still not correct.

It may also be argued that electron density in $\pi - \pi^*$ transitions is mainly above and below the plane of the base, approximately where $p$ orbital density is greatest. Calculations were done in which each
in-plane monopole was placed 0.75 Å above and below the atom on a line perpendicular to the plane of the base, as was done for the out-of-plane background oscillator. The value of the monopole at each of these positions was assigned half the value associated with the corresponding atom. This procedure produced only relatively insignificant increases or decreases in magnitude.

(D) Shifting the Monomer Absorption Bands

For the dimer case, at least, it becomes apparent that no obvious alternative exists for correcting the positions of the maximum and minimum at 215 and 202 nm than to move the position of the 207 nm monomer absorption band from which these two polymer bands are derived. When the Re α and Im α for the 207 band were shifted 0.8 kK to the red, putting the absorption maximum at about 210.5 nm, the new polymer maximum and minimum became 6.8 at 219 nm and -15.1 at 205 nm, in much better agreement with measured values. The calculated CD above 230 nm did not change significantly. ApA in B-DNA geometry was similarly improved near 220 nm, but below 210 the magnitude of all CD bands became much larger. Small errors in the long wavelength band positions were found to be correctable by moving the 260 oscillator a few nanometers blue.

Red shifting the 207 nm monomer band does not help in the case of oligo A_{20} but merely moves the
negative band below 220 nm further to the red. Matrix method calculations in which various oscillators are omitted show that the order in which large positive or negative $R_K$ bands near 207 nm appear as well as their relative strengths are strongly influenced by interactions with oscillators at other frequencies. For instance, CD calculations done with only the 260, 240 and 207 nm oscillators produced a large positive band at 211 nm and a large negative band at 200 nm. Introducing either the background oscillators or the 187.5 nm oscillator alone reverted the CD near 210 nm to the pattern observed in the full calculation.

As mentioned above, a different choice of monopoles might correct the sign of the CD below 230 nm, but so far none has been found. Another approach that showed promise was exchanging the parameters (monopoles, $\theta$, $R_i$) for the 207 and 187.5 nm oscillators. In $A_{20}$, a small positive peak appeared at 214 nm, and red shifting the 207 nm monomer band did indeed produce a better position for this band. However, agreement in the 230-250 nm region worsened as did the overall ApA calculation done with identical parameters.

VI. Discussion

Polarizability theory, using $\varepsilon_{\text{diel}}$ equal to 2 and the proper choice of monomer parameters, predicts the optical properties of ApA and ribo oligo A throughout the measurable UV wavelength region reasonably well.
Correct predictions of the basic CD pattern for these simple polymers, above 230 nm, are not unique to polarizability theory, however. The question to be answered is to what extent, or in what cases, will polarizability theory give substantially better or different results than simpler theories.

In Chapter 1, equations for the CD of a two oscillator system were determined analytically by all three theories. Although the necessity of arbitrarily assigning polymer band shapes in first order perturbation theory and in the matrix method precludes direct comparison with polarizability theory, polarizability theory results are expected to differ very little from the other theories unless $G_{12}$ or the monomer absorption is exceptionally large. Each of the 7 oscillators in ApA tends to couple mainly with itself in a quasi-two oscillator system, so it is not surprising that all three theories give very similar results above 230 nm. Changing $\varepsilon_{\text{dieel}}$ from 1 to 2 in either the polarizability theory or matrix method calculation of ApA approximately, though not exactly, cuts the magnitude in half. For ApA, higher order terms are clearly not very important even with 7 oscillators per base.

This simple relationship of magnitude to dielectric constant is rarely the case for larger polymers. It is apparent from the non-linear effect of $\varepsilon_{\text{dieel}}$ on oligo A$_{20}$ CD that higher order terms can be
important even in a very simple sequence. Increasing $\varepsilon_{\text{diel}}$ to 2 reduces the 265 nm peak maximum by only 25%, not by a factor of two, and measurably shifts $\lambda_{\text{max}}$. At $\varepsilon_{\text{diel}} = 1$, the polymer bands derived from the adenine 207 nm band retain the pattern observed in the ApA spectrum. At $\varepsilon_{\text{diel}} = 2$, the $G_{ij}$ term between adjacent 207 oscillators, small to begin with, becomes so small that interference from strong interactions with other oscillators totally alters the CD around 210 nm. One reason for the nonproportionality of large polymers to $\varepsilon_{\text{diel}}$ seems to be due to the large number of polymer bands per monomer oscillator, evident in the matrix method calculation shown in figure (4). The few polymer bands giving rise to large $R_K$ values are normally spaced much closer together than in the dimer, particularly when $G_{ij}$ values involved are small. Since strong bands in closely spaced groups sometimes cancel weaker ones, the final CD is highly sensitive to the splitting and to the relative strength of each band, both of which are presumably influenced by higher order $G_{ij}$ terms. Regardless of the mechanism, theories carried only to first order in $G_{ij}$ and $V_{ij}$ will apparently predict CD band positions, magnitudes, and shapes for large polymers occasionally quite unlike and less accurate than spectra predicted by higher order theories. The polarizability theory calculation of oligo $A_{20}$, even using $\varepsilon_{\text{diel}} = 1$, is in
much better agreement with the measured CD of poly rA than is the first order perturbation calculation of poly A in RNA geometry. The latter calculates the first CD peak and trough to be at 270 nm and 245 nm, respectively, with magnitudes of +67.0 and -16.5. (These magnitudes should actually be multiplied by 2, due to an error found in the first order perturbation theory after publication.)

The matrix method, all order in $V_{ij}$, is capable of calculating CD spectra remarkably similar to the CD spectra calculated by polarizability theory when correct band shapes and widths are chosen. For some purposes, particularly when only a rough idea of the CD pattern is desired, the matrix method is more convenient. It is limited in the total number of oscillators the program may handle in the same way as is the polarizability theory program. However, computer time required for large polymer calculations is significantly shorter, because once the matrix is diagonalized, any number of frequency points can be calculated with relatively little addition of computer time. The polarizability theory program must handle a unique matrix at each frequency point, and hence calculations involving many oscillators and a large number of frequency points are prohibitive. However, the matrix method may on occasion be deficient in predicting the exact shape of spectra, as evidenced by its failure to
calculate a shoulder at 230 nm for oligo $A_{20}$ that is predicted in the polarizability theory calculation.

In the matrix method, two non-identical transitions will interact as delta functions separated by $\nu_{1\text{ max}} - \nu_{2\text{ max}}$. The same interaction in polarizability theory is frequency dependent, since the theory allows portions of each transition to interact with a smaller frequency separation. It is this frequency dependence of interaction which leads to higher order terms that do not appear in the matrix method equation. It is therefore not surprising that the two theories are more divergent when an effective dielectric constant of 1 rather than 2 is used, i.e. when $G_{ij}$ values are twice as large.

In CD calculations of other dimers, not shown here, significant differences in spectra predicted by the two theories were often observed in peak/trough ratios and occasionally in band shapes. The differences were most pronounced in certain heterogeneous dimers such as UpC, CpA, and GpU where the calculated CD was largely all positive or all negative over large regions of the spectrum, i.e., where non-conservative effects were predominant.

Aside from comparison of theoretical methods, the results of this work definitely indicate that the introduction of an effective dielectric constant will generally produce beneficial and sometimes important
improvements in the calculated spectra. Admittedly, the practice of reducing all $G_{ij}$ values uniformly is oversimplified, but it will probably suffice until more rigorous treatments of the problem are explored. The best results are obtained when $\varepsilon_{\text{die}}$ is set at 2. Larger integer values tend to reduce the CD magnitude too much and do not significantly improve shape of the spectra. A dielectric constant of 2 applied to coupling of transitions separated by and in the environment of other nucleic acid bases seems quite reasonable since $\varepsilon_{\text{die}}$ values for many organic solids fall between 2 and 3.

On the other hand, the use of split monopoles, above and below the plane of the base, has a relatively small effect in both ApA and oligo A calculations. Apparently, the bases are sufficiently separated such that placing half the charge closer together and half farther apart does not radically change the value of $G_{ij}$. The smallest effect is observed above 230 nm where the major transition generally has monopoles highly localized on two atoms. Where the monopoles are more spread out, overlap of monopoles on adjacent bases will be greater and the effect of split monopoles somewhat more pronounced. Nevertheless, the computer time saved by using all in-plane monopoles for in-plane transitions outweighs the small error that results. At the least, trial calculations testing
variations in parameters for large changes in CD can be done this way before doing more precise calculations with split monopoles.

The actual choice of monopoles may be one of the most critical factors in predicting correct CD patterns. At this point, the only reliable method of selecting monopoles, as with all monomer properties, is to search for the set most often successful in calculations. Oligo A calculations alone do not provide a very thorough test of the CNDO monopoles. It does appear, however, that when used they should not be adjusted for $\theta$. The CNDO method probably correctly predicts the general location of transition charge density, but it may not give quite accurate relative sizes for the monopoles. Adjusting the monopoles as described above tends to spread the transition out over the entire base.

In general, the combination of CNDO monopoles and the available experimental data for the near UV transitions of adenine works reasonably well in predicting optical properties near 260 nm. Nevertheless, the use of monomer data collected from monomers in environments other than the one being calculated must be questioned. Transition directions are measured in crystals or films, and absorption bands are resolved from spectra measured in aqueous solution. Absorption measurements of monomers in non-aqueous solvents
and in vacuum \(^{15}\) suggest that the electronic transitions may shift and/or change in strength when the base is placed in the non-aqueous environment of other bases.

The variations reported here in the number, frequency location, and transition angles of oscillators were not meant to constitute a comprehensive study of empirical adjustments to monomer properties, but were done only to give some indication of the types of spectral changes these particular parameter changes could be expected to produce. This method of approaching errors in the calculated spectra is designed to narrow the possible range of empirical adjustments that might prove beneficial. For example, there appears to be no alternative for correcting the ApA CD calculations but to assume a red shift in the monomer 207 nm transition. Further corrections in the 200 nm region may very well necessitate the addition of a third oscillator there. A change in the position, intensity, or monopoles of the 240 nm adenine transition might reduce the magnitude of the negative band at 250 nm in the oligo A\(_{20}\) calculation while making only minor changes in the ApA calculation.

Unfortunately, it is difficult to make conclusive statements concerning monomer properties when remaining discrepancies between measured and calculated CD may be due to other factors. The assumed geometry is
the most obvious source of error. The temperature de-
pendence of the measured CD has already been discussed,
but even at very low temperatures, the most stable con-
figuration may differ from strict RNA or B-DNA in the
absence of constraining hydrogen bonds to another
strand. It is possible that the chosen monomer para-
eters are already sufficiently correct for the oligo
$A_N$ calculations, for instance, and that the observed
CD in the 220 nm region arises from poly A having a
geometry somewhere between RNA and B-DNA. Calculations
need to be done on polymers for which the geometry is
better defined in order to accurately adjust para-
meters or to estimate the importance of such tradi-
tionally neglected problems as direct solvent effects
and $n - \pi^*$ transitions.

In summary, the strength of polarizability the-
ory, aside from being all order in $G_{ij}$, is in its
incorporation of the frequency dependence of monomer
absorption bands. First order perturbation theory is
quite limited by its formalism in the type of band
shapes it can produce. It is not expected to be very
successful where, for instance, a strong band at 260
nm interacts with one at 280 nm, or in cases where
higher order terms are important. The matrix method
will generally give $R_K$ patterns that resemble polar-
izability theory results, but CD spectra generated
from the $R_K$ patterns will sometimes differ on usually
small but occasionally important details. Unless $R_K$ patterns are specifically desired or computer time is a serious problem, polarizability theory will avoid the problem of defining polymer band shapes and will produce more accurate spectra.
CHAPTER 3 REFERENCES


8. Davis, R.C. and Tinoco, I., Jr. (1968) Biopolymers 6, 223-242. (See ref. 4 for transform.)


Chapter 4
CALCULATIONS FOR DOUBLE STRANDED
POLYNUCLEOTIDES OF REPEATING SEQUENCE

I. Objective

The availability of measured CD spectra\(^1-^9\) for
double stranded polynucleotides of simple repeating
sequences probably provides the best opportunity for
evaluating the reliability of polarizability theory
and the chosen set of monomer properties. The simple
sequences are preferable to random sequences for this
purpose because they have highly sequence specific
CD spectra of larger magnitude. This is almost cer-
tainly a result of having simpler rotational strength
patterns of large and relatively few bands, giving
rise to less cancellation and more distinctive spectra.

The structure of poly rA:rU has been determined by
X-ray fiber diffraction\(^10\) and found to have standard
A-RNA geometry, the conformation observed for fibers
of double stranded viral RNA. Poly rI:rC (I = inosine,
an analog of guanine) studied under identical condi-
tions was also in A-RNA geometry. No evidence yet
exists that any of the double stranded ribo polynucleo-
tides have significantly different structures in solu-
tion. Therefore, calculations done on simple sequences
in RNA geometry will be considered the best test of
theory and parameters. The four sequences that will
be discussed at length in this chapter are poly A:U (all adenine on one strand and all uracil on the other), poly AU:AU (strictly alternating A,U on each strand), poly G:C, and poly GC:GC.

Deoxy polynucleotides, on the other hand, give much indication of structural variation. DNA CD spectra are solvent and salt dependent, and X-ray work has produced considerable evidence, albeit conflicting, for sequence specific structures that are not standard B or C form DNA. Hopefully, as progress is made toward correctly calculating the CD of the ribo polymers, CD calculations will provide additional evidence, positive or negative, concerning the validity of proposed DNA solution structures. Calculations for all four sequences in B-DNA and C-DNA geometry as well as some of the proposed non-standard geometries will be presented here, but conclusions can be drawn from such calculations at this point only very tentatively.

II. RNA Calculations

Figure (8) shows both the calculated and measured CD spectra for poly A:U, poly AU:AU, poly G:C, and poly GC:GC. Calculations were all done with polarizability theory, the monomer parameters listed in Chapter 2, and an effective dielectric constant of 2. The calculations shown are for 10 base pairs; nearly 1 turn of the RNA helix, for each of the sequences. For
Figure 8. Calculated CD spectra for 10 base pair poly A:U, poly AU:AU, poly G:C, and poly GC:GC in RNA geometry ———. Measured CD spectra for poly rA:rU, poly rAU:rAU, poly rG:rC, and poly rGC:rGC ———. All calculations were done with polarizability theory and $\varepsilon_{\text{diele}} = 2$. 
simplicity, "poly" rather than "oligo" will be used in reference to calculated spectra for the remainder of this discussion. The measured spectra were not corrected for monomer CD, as discussed in Chapter 1. Occasionally somewhat better agreement was obtained by subtracting the relevant monomer CD, but the improvement was not sufficient to outweigh the arguments against this procedure.

The calculated CD spectrum of 10 base pair A:U agrees extremely well in sign and band position with the measured spectrum of poly rA:rU. The calculated magnitude is not quite correct, particularly at 265 nm where the calculated maximum is about two times too large. Nevertheless, this should be considered good agreement. Oligo adenylic acid calculations, discussed in Chapter 3, showed the relative magnitudes of peaks and troughs in the 260 nm region to be sensitive to transitions in the 200 nm region and in the far UV, transitions whose properties can only be approximated. The calculation is considerably better in both magnitude and band position than the spectrum calculated by first order perturbation theory.²⁸

Unfortunately, agreement in band shape is less satisfactory for the poly AU:AU calculation. The calculated spectra for poly A:U and poly AU:AU are, in fact, almost identical above 240 nm. This is not surprising, since the main contributions to the CD at
long wavelength come from the adenine 260 nm and uracil 262 nm oscillators. These two transitions have nearly parallel dipole moments and have been assigned very similar transition monopoles. The shape of the measured poly AU:AU CD spectrum suggests that there might be hidden negative rotational strength at long wavelength (missing in the calculated CD) that reduces the CD magnitude near 270 nm and blue shifts $\lambda_{\text{max}}$.

The calculated spectrum of 10 base pair G:C is in reasonable agreement with the measured CD. It is at least much better in magnitude and somewhat better in shape than earlier calculations.\textsuperscript{28} The calculation produces CD bands of proper sign and approximately correct position but predicts incorrect magnitudes. The calculation may be failing to predict negative rotational strength in the 280-290 nm region, which would correct the magnitude in this region as well as blue shift the 275 nm maximum toward a more accurate position.

Virtually no agreement with measurement is reached in the poly GC:GC calculation. First order perturbation calculations\textsuperscript{28} were equally unsuccessful.

Calculated absorption spectra for these four sequences are shown in Figure (9). Measured absorption spectra have shapes very similar to the calculated spectra for poly A:U and poly AU:AU but have different $\varepsilon_{\text{max}}$ and $\lambda_{\text{max}}$ values. Measured maximum extinction co-
Figure 9. Calculated absorption spectra for four 10-base pair polynucleotides in RNA geometry: poly A:U and poly G:C ——; poly AU:AU and poly GC:GC ———.
Efficiencies are $6.9 \times 10^3$ for poly A:U$^{1,29}$ and $6.1 \times 10^3$ for poly AU:AU$^{1,30}$ both at about 260 nm. The $\lambda_{\text{max}}$ for poly AU:AU appears to be slightly to the red of $\lambda_{\text{max}}$ for poly A:U, as predicted by the calculations, but the calculated absorption maxima are both too large and about 5 nm too far to the blue. The calculated absorption for poly G:C is in best agreement with measurement, both in shape and magnitude. The measured spectrum$^{31}$ has a small shoulder at 280 nm, peaks at about $\varepsilon = 7.5 \times 10^3$ at 260 nm, and has a minimum of $4 \times 10^3$ at 230 nm. Agreement for poly GC:GC, however, is poor. The measured spectrum$^{32}$ resembles the measured absorption of poly G:C in shape, with a larger shoulder at 270 nm and a maximum of $6.6 \times 10^3$ at 260 nm. The calculation overestimates the magnitude at 270 nm and underestimates it at 250 nm.

Agreement or disagreement between measured and calculated absorption near 260 nm can not be taken as proof or disproof of the reliability of 260 nm region monomer properties. Calculations in which the 200 nm or background oscillator parameters were varied have shown the calculated absorption near 260 to be much more sensitive to distant oscillators than is the calculated CD. As discussed in Chapter 3, this is to be expected, since hypo- or hyperchromism is entirely dependent on borrowing from other absorption bands.

Arnott et al.$^{10}$ report that upon increasing salt
concentration, poly A:U and poly I:C undergo a transition from 11 fold A-RNA to a very similar but 12 fold helical structure, A'-RNA. CD calculations done with A'-RNA geometry were almost identical to A-RNA calculations. Certainly no conclusions could be drawn from the calculations as to which is the solution structure of RNA.

III. B-DNA Calculations

Calculations for the same four sequences in B-DNA geometry are shown in Figure (10). In this case, calculations were done with 10 base pairs per polymer, i.e. one turn of the B-DNA helix. Uracil monomer data, with slight alterations in $\theta$ for the 262 and 175 nm oscillators (see Table (1)), was used for thymine in all DNA calculations.

The calculated CD spectra for both poly A:T and poly AT:AT in B-DNA geometry are very similar to the RNA calculations (see comparison in Figure (11)). The magnitude of the first maximum is reduced in the B-DNA calculation, in agreement with observation. However, further agreement with measured spectra is very poor for poly A:T, and the calculated poly AT:AT spectrum suffers from the same inaccurately predicted $\lambda_{max}$ seen in the RNA calculation.

Perhaps the most successful calculation is poly G:C. The calculated peak at 250 nm should be further
Figure 10. Calculated CD spectra for 10 base pair poly A:U, poly AU:AU, poly G:C, and poly GC:GC in B-DNA geometry ———. Measured CD spectra for poly dA:dT $^4$, poly dAT:dAT $^5,9$, poly dG:dC $^5,9$, and poly dGC:dGC $^4$ ———.
Poly dA·dT

Poly dG·dC

Poly dAT·dAT

Poly dGC·dGC

Wavelength (nm)
Figure 11. Calculated CD spectra for poly A:U, poly AU:AU, poly G:C, and poly GC:GC in A-RNA (A) geometry -----; in B-DNA (B) geometry --- ---; and in C-DNA geometry ······.
to the red, and the calculation at short wavelength is clearly inadequate, but the small positive peak at 283 nm is very accurately predicted, and the overall magnitude is in excellent agreement with the measured spectrum. This is in sharp contrast to the first order perturbation theory calculation,\textsuperscript{28} which predicts essentially the same $\lambda_{\text{max}}$ and $\lambda_{\text{min}}$ for poly G:C in RNA and B-DNA geometries.

The calculated CD of poly GC:GC is again similar to the RNA calculation. Measured spectra\textsuperscript{4-6} for poly dGC:dGC differ depending on whether the measurements are made at low (0.2 M NaCl) or high (20% W/W NaCl) salt concentrations. A laser Raman scattering study\textsuperscript{33} indicated that the low salt form is probably a B-form structure, whereas the high salt form does not resemble known conformations. The calculated CD is closer to the spectrum measured at low salt, but agreement even in this case is generally poor. Only the CD near 285-290 nm, i.e. CD arising from the guanine 281 oscillator, is correctly positioned.

Despite failures to predict measured CD, it is interesting to note that polarizability theory is capable of occasionally predicting quite different band shapes for the same sequence in RNA versus B-DNA geometries, for instance in poly G:C. This has not been the case for first order perturbation theory calculations.
IV. **C-DNA Calculations**

Calculated spectra for the four sequences in C-DNA geometry are shown in Figure (11), along with the same sequences calculated in RNA and B-DNA geometries. Assuming the calculations to be approximately valid, differences between CD spectra for polynucleotides in these conformations are somewhat sequence dependent. Calculated C-form spectra are generally B-like except at long wavelength, where a loss of positive rotational strength is observed for A/U polymers and an increase in negative rotational strength is observed for G/C polymers. B-form spectra are generally slightly reduced in magnitude above 240 nm in comparison to RNA spectra except for poly G:C, where the calculated shape is significantly different.

Evidence for the existence of C-form DNA in solution is based on similarities between the CD spectra of natural DNA measured under special solution conditions, e.g. high LiCl concentration\(^ {17,18} \) or in ethylene glycol,\(^ {11,12} \) and the CD spectra of Li-DNA films at 75% humidity,\(^ {15} \) conditions which give C-form X-ray patterns.\(^ {34} \) The calculated C-form spectra for poly AU:AU does not at all agree with the CD measured for poly dAT:dAT in 5.6 M LiCl\(^ {17} \). There is, however, reasonable agreement between calculated C-form spectra of the two A/U polymers and room temperature measurements of these molecules in ethylene glycol.\(^ {12} \)
V. Parallel and Perpendicular Components

Flow orientation methods are now available for partially aligning long helical molecules in solution. Independent measurements can then be made of the optical properties for light incident parallel and perpendicular to the helix axis. Measurements made on natural, random sequence RNA in buffer and on DNA in buffer (B-DNA) and in 6M LiCl or in 95% ethylene glycol (perhaps C-form DNA) show the CD components to be much more sensitive to structure than is the average, isotropic CD. Figures (12 a-f) show currently available measured component CD spectra.

The calculated CD component spectra are shown, in Figure (13), for poly A:U and poly G:C in RNA, B-DNA, and C-DNA geometries. The component labeled parallel is the CD calculated for light incident along the z axis, where the z axis is defined to be the helix axis. The perpendicular component is the average of the CD for light incident along the x axis and along the y axis.

Measured component data is not yet available for these sequences. The measured isotropic CD of poly I:C is similar, however, to that of poly G:C above 260 nm. In this region, the measured spectrum of poly I:C non-shows a conservative perpendicular component and a conservative parallel component that are roughly the opposites of the calculated components for poly
Figure 12. Measured\textsuperscript{35,36} component and isotropic CD spectra for naturally occurring RNA and DNA:

\[(\varepsilon_L - \varepsilon_R)_{xx} + (\varepsilon_L - \varepsilon_R)_{yy}, \text{ i.e. } 2 \times \text{CD for light incident perpendicular to the helix axis}\]

\[\cdots \cdots \cdots \text{to the helix axis} -o-o-o-;
\]

CD incident parallel to the helix axis, \((\varepsilon_L - \varepsilon_R)_{zz}, \cdots \cdots \cdots\); and isotropic CD \(\cdots \cdots \cdots\). (a) RNA in 0.1 M NaF and 0.01 M PO\(_4\) buffer, pH 7.0; (b) DNA in 0.01 M NaF and 0.01 M PO\(_4\) buffer, pH 7.0; (c) DNA in ethylene glycol; (d) DNA in 6 M LiCl. Also shown are similar spectra for (e) poly U:A:U and (f) poly I:C. These spectra were kindly made available to us by S.Y. Chung and G. Holzwarth.
(a) OBSERVED CD, DOUBLE-STRANDED RNA

\[ [\varepsilon_L - \varepsilon_R]_{33} \]

\[ [\varepsilon_L - \varepsilon_R] \]

\[ [\varepsilon_L - \varepsilon_R]_{11} + [\varepsilon_L - \varepsilon_R]_{22} \]

WAVELENGTH (nm)
(b) OBSERVED CD, DNA IN BUFFER
Figure (c) shows the observed CD, DNA in Ethylene Glycol.

The graph presents the wavelength (nm) on the x-axis and the difference in molar ellipticity \( \epsilon_L - \epsilon_R \) on the y-axis. The graph highlights the following transitions:

- \( [\epsilon_L - \epsilon_R]_{33} \)
- \( [\epsilon_L - \epsilon_R] \)
- \( [\epsilon_L - \epsilon_R]_{11} \)
- \( +[\epsilon_L - \epsilon_R]_{22} \)
OBSERVED CD, DNA IN 6M LiCl

[Diagram showing CD spectra with labeled peaks: $[\varepsilon_L - \varepsilon_R]_{33}$, $[\varepsilon_L - \varepsilon_R]_{II}$, $+[\varepsilon_L - \varepsilon_R]_{22}$]
OBSERVED CD, POLY(A)·2 POLY(U)

\[
[e_L - e_R]_{33}
\]

\[
[e_L - e_R]_{22} + [e_L - e_R]_{11}
\]
OBSERVED CD, POLY(I)·POLY(C)
**Figure 13.** Calculated CD spectra for poly A:U and poly G:C showing the isotropic CD ——; CD for light incident parallel to the helix axis ---; and CD for light incident perpendicular to the helix axis ·····: (a) in RNA geometry, (b) in B-DNA geometry, and (c) in C-DNA geometry. The isotropic CD is equal to 1/3 (parallel CD + 2 x perpendicular CD).
G:C. Calculated versus measured relative magnitudes of the two components do not correspond, however. The measured components of poly U:A:U (a triple stranded polymer with two strands of non-alternating uracil and one strand of adenine) bear no resemblance at all to the calculated components of poly A:U. This might only reflect the difference in number of strands or may indicate serious trouble in the poly A:U calculations. Interestingly, the calculated components of poly A:U in the three geometries agree rather well in overall shape (conservative versus non-conservative, relative magnitudes, etc.) with the measured components of natural RNA and DNA in buffer, and, to a lesser extent, with the high salt measurement of DNA.

VI. Effect of Dielectric Constant

The conclusions drawn in Chapter 3 concerning the use of an effective dielectric constant are supported by calculations on the double stranded polynucleotides. Without exception, agreement with measured spectra was much better using a dielectric constant of 2 rather than 1, in calculations where any agreement at all was evident.

Figure (14) shows the calculated spectra of poly A:U in RNA geometry with $\varepsilon_{\text{die}} = 1, 2, \text{and} 4$. The corresponding spectra for poly AU:AU are similar, except that the change from $\varepsilon_{\text{die}} = 1$ to 2 produced an increase
Figure 14. Calculated CD spectra for (a) poly A:U and (b) poly G:C in RNA geometry using $\varepsilon_{\text{die}} = 1$ -- - - -; $\varepsilon_{\text{die}} = 2$ ---; and $\varepsilon_{\text{die}} = 4$ ······.
rather than decrease in CD magnitude near 265 nm only in the poly A:U calculation. This increase in magnitude was small and did not much affect agreement with measurement, while the red shift that resulted brought the calculated \( \lambda_{\text{max}} \) directly in line with the measured \( \lambda_{\text{max}} \). In DNA geometry, the effect of higher order terms in \( G_{ij} \) appeared to be smaller. For both poly A:U and poly AU:AU, doubling the value of \( \varepsilon_{\text{die}} \) reduced the magnitude of the CD above 240 nm approximately by half and caused only small shifts in the 270 nm band.

The calculations for polymers of guanine/cytosine composition were much more sensitive to \( \varepsilon_{\text{die}} \), for example poly G:C in RNA geometry, also shown in Figure (14). The B-DNA calculations were generally more proportional to \( G_{ij} \) than the RNA calculations, but for poly G:C and poly GC:GC in B-DNA geometry, the \( \varepsilon_{\text{die}} \) change from 1 to 2 increased the magnitude of the negative bands at 227 and 267 nm, respectively. In the case of poly G:C, use of \( \varepsilon_{\text{die}} = 2 \) brought the calculated magnitude into nearly exact agreement with the measured magnitude.

VII. Calculations Done with Non-Standard Geometries

Recent X-ray diffraction work done by Arnott et al. suggests that deoxy polynucleotides in which purine and pyrimidine nucleotides alternate along each
strand may adopt an unusual conformation in solution. In particular, X-ray patterns for poly dAT:dAT and poly dGC:dGC are consistent with a right handed, 8 fold helix with an axial rise of 3.03 Å per residue. These parameters result in a structure of unusually dense packing.

Calculations for poly AT:AT and poly GC:GC using the coordinates provided by Arnott et al. are shown in Figure (15). For poly AT:AT, the calculated spectrum is quite different from the regular B-DNA calculation, but it is in much worse agreement with the measured spectrum of poly dAT:dAT. For poly GC:GC, the negative CD calculated at long wavelengths is in better agreement with the CD measured at high salt rather than low salt, but calculated band positions actually agree with neither.

Arnott et al. have also published X-ray derived coordinates for poly dA:dT and for the triple strand polymer poly dT:dA:dT. The coordinates for poly dA:dT are very similar to standard B-DNA coordinates, so it was not surprising to find that the CD calculated with these coordinates was essentially indistinguishable from the calculation of poly A:T shown in Figure (11). The X-ray coordinates for poly dT:dA:dT gave a calculated CD spectrum also very similar to the B-DNA calculation of poly A:T, but with much lower magnitude, e.g. with $\varepsilon_L - \varepsilon_R = 4.2$ at 264 nm and -2.6 at 247.5 nm.
Figure 15. Calculated CD spectra for poly AT:AT and poly GC:GC using the X-ray coordinates of Arnott et al.25
The authors point out that these coordinates are conformationally similar to X-ray coordinates for poly rU:rA:rU. The measured CD of the triple strand ribopolymer\(^2\) is very similar to the measured CD of poly rA:rU, so the calculation for poly dT:dA:dT is reasonable.

VIII. Empirical Search for Better Monomer Parameters

(A) Monopoles

As pointed out in the discussion of ApA and oligo\(A_N\) calculations, no easy method is available for systematically varying monopoles to determine the effect of monopole distribution on the calculated CD. In those calculations, the direct use of CNDO monopoles for the first two transitions of adenine was found to give better results than the use of CNDO monopoles adjusted to give experimentally measured \(\theta\) values. The same adjustment of monopoles was done for the first two oscillators of all four bases and tested in both RNA and B-DNA calculations of the four sequences. Results were not significantly different. Somewhat better agreement was attained using adjusted monopoles only in the case of poly G:C in B-DNA geometry. Overall, the original CNDO monopoles gave equally good, and sometimes better, results than the adjusted monopoles.

The reliability of using CNDO monopoles, directly
as calculated by the CNDO method, should not be overestimated, however. Good agreement with measured CD spectra was obtained most consistently by reversing the assignment of guanine monopoles for the first two transitions of guanine. In the terminology of ref. (37), the CNDO method predicts that monopole set II should be assigned to the lowest energy transition and set I to the next higher energy transition. In contrast, best results were obtained using set I for the 281 nm oscillator and set II for the 252 nm oscillator. The only exception, in which the opposite order worked better, was the CD calculation of poly GC:GC in RNA geometry. In this case, the longest wavelength CD band was negative, in accordance with measurement. The remainder of the calculated spectrum was in better agreement with measurement but not clearly accurate. Using the same monopole assignment for guanine in poly G:C, however, gave no agreement at all in either the RNA or the B-DNA calculation.

Calculations were also done for the four sequences, in RNA geometry, placing the monopoles for all oscillators 0.75 Å above and below the plane of the base, as described in Chapter 3. In the double stranded simple sequence polymers as well as in single stranded oligo A, this procedure produced no significant differences in the calculated CD patterns. The calculated CD of poly G:C in RNA geometry showed the largest change;
the CD band at 271 nm shifted to 275 nm and decreased in magnitude to $\varepsilon_L - \varepsilon_R = 6.1$, while the small band at 251 nm was reduced to a shoulder along the $\varepsilon_L - \varepsilon_R = 0$ line. The negative CD at short wavelength increased in magnitude considerably.

(B) Transition Angles

Variation in the transition angle $\theta$ for any given oscillator is expected to have a larger effect on the calculated CD of a double stranded heteropolymer, particularly one of alternating sequence such as poly AU:AU, than on the CD of a single stranded homopolymer such as poly A. If $\theta$ for the 260 nm oscillator of adenine, for instance, is changed from its present value of 177°, the adenine $\mu_i$ will no longer have a direction approximately parallel to the transition dipole of the uracil 262 nm oscillator. Consider the $R_{ij} \cdot E_i \times E_j$ term in eq. (17), in regard to coupling of the 260 nm transitions only. A change in $\theta$ for the 260 nm oscillator of adenine will not change any of these terms in poly A but will give considerably different terms for nearest neighbor interactions between A and U in poly AU:AU.

PROGRAM ROTOPR was used to test systematic variations in $\theta$ only for the 260, 240, and 207 nm oscillators of adenine in RNA geometry calculations of poly A:U and poly AU:AU. Changes in $\theta$ for the 240 and 207 nm
oscillators had very little effect on the calculated CD near 260 nm. Figure (16) shows the CD and absorption for poly AU:AU as θ for the 260 nm oscillator is varied from −60°, through −135° (the value of θ calculated from the 260 nm oscillator monopoles), to +150°. Interestingly, the value of −135° gave the best agreement in magnitude for both polymers, but it did not cause a large enough blue shift in the first poly AU:AU CD maximum to give agreement that would be considered substantially better.

In light of the above result, calculations were done for all four sequences, in both RNA and B-DNA geometries, using θ values for the first two transitions of each base calculated from the corresponding CNDO monopoles. Results for RNA calculations are shown in Figure (17). Agreement with measured CD was occasionally improved, particularly the magnitude of poly A:U at 260 nm and the shape of the poly G:C CD spectrum at 280 nm. Unfortunately, the B-DNA calculations were not at all improved. The poly A:U results were again the same as the RNA calculation. Poly G:C and poly GC:GC agreement became slightly worse, and the calculated CD of poly AU:AU became totally negative above 235 nm. Calculations for deoxy polymers in non-standard geometries also remained essentially the same or became worse.
Figure 16. Calculated CD and absorption spectra for poly AU:AU in RNA geometry using standard parameters for all adenine oscillators except $\theta = 120^\circ$ for the 240 nm oscillator and the following values of $\theta$ for the 260 nm oscillator: (1) $-60^\circ$; (2) $-90^\circ$; (3) $-120^\circ$; (4) $-150^\circ$; (5) $180^\circ$; (6) $150^\circ$. 
Figure 17. Calculated CD spectra for (a) poly A:U and poly AU:AU, and (b) poly G:C and poly GC:GC in RNA geometry using values of $\theta$ for the two longest wavelength oscillators of each base calculated from the corresponding monopoles by eq. (36).
\begin{itemize}
\item (b) Poly GC:GC / Poly G:C
\end{itemize}
(C) Shifting the Monomer Absorption Bands

Neither our resolution of adenosine monophosphate nor our resolution of uridine monophosphate can be considered indisputable. Some evidence exists for a small absorption band on the red side of 260 nm in adenine (see ref. (38) for review of evidence) and for a small absorption band near 250 nm in uracil.\textsuperscript{39} Both of these possibilities were tested to see if they might give better calculated spectra for poly A:U and poly AU:AU in RNA geometry. Uracil was resolved into two bands, one of $\varepsilon_{\text{max}} = 8.3 \times 10^3$ at 265.5 nm and a smaller band of $\varepsilon_{\text{max}} = 5 \times 10^3$ at 248 nm. The new values of $\theta$, taken from ref. (39), were $9^\circ$ and $-35^\circ$, respectively. The first two sets of CNDO monopoles were used for these oscillators, and new monopoles were calculated for the 207 nm oscillator using PROGRAM BASES. For both polymers, the calculated CD spectra were nearly the same as calculated with the original set of parameters for uracil. The first peak and trough in each spectrum were blue shifted by about 2 nm with little change in magnitude.

For adenine, the 240 nm oscillator was dropped and a new resolution used in which a small band of $\varepsilon_{\text{max}} = 4 \times 10^3$ was placed at 270 nm and a larger band of $\varepsilon_{\text{max}} = 13.6 \times 10^3$ was placed at 257 nm. The first two sets of CNDO monopoles were assigned to the new 270 and 257 nm transitions, both in the same order as
before and in reverse order. Variations in $\theta$ for both oscillators were tested as well. Only one combination of parameters gave good results for both poly A:U and poly AU:AU; the 270 nm band assigned monopole set I (originally assigned to the 260 nm oscillator) and $\theta = \pm 135^\circ$, and the new 257 nm oscillator assigned monopole set II (originally assigned to the 240 nm oscillator) and $\theta = 120^\circ$. Figure (18) shows the CD spectra calculated with these parameters. The calculated absorption maxima were $8.4 \times 10^3$ at 260 nm for poly AU:AU and $9.6 \times 10^3$ at 255 nm for poly A:U. The calculated CD spectra are in excellent agreement with measured CD. Unfortunately, the B-DNA calculations using these parameters were very similar to the RNA calculations, the calculated CD of poly A:U showing no more agreement with the measured CD of poly dA:dT than the original calculation. Furthermore, ApA and oligo A CD spectra calculated with these parameters in RNA geometry were totally wrong, the CD in both cases being entirely negative above 225 nm. The calculated CD of oligo A$_{20}$ in B-DNA geometry resembled the shape of the original calculated spectrum, but band positions and magnitudes no longer matched the measurements for either poly rA or poly dA. Finally, the calculations using the recent X-ray coordinates
for poly dAT:dAT and poly dT:dA:dT changed but did not improve in agreement.

IX. Discussion

The success of polarizability theory and the cho-
Figure 18. Calculated CD spectra for poly A:U and poly AU:AU in RNA geometry using alternate parameters for the monomer properties of adenine between 230 and 280 nm; oscillator 1 at 270 nm with $\epsilon_{\text{max}} = 4 \times 10^3$, monopole set I, and $\theta = -135^\circ$; oscillator 2 at 257 nm with $\epsilon_{\text{max}} = 13.6 \times 10^3$, monopole set II, and $\theta = 120^\circ$. 
sen set of monomer properties in predicting the CD spectra of repeating sequence ribo polynucleotides is remarkable in some cases and disappointing in others. The accuracy of some calculations, in location of maxima and minima and particularly in predicting absolute magnitude, definitely suggests that polarizability theory coupled with the use of an effective dielectric constant should be considered adequate and sound. Remaining discrepancies between calculated and measured CD spectra are very likely due either to misjudgement of the geometry or errors in the choice of monomer parameters. It is conceivable, despite lack of evidence, that ribo polynucleotides do not always exist in standard RNA geometry in solution. Alternating sequences are particularly suspect because of their self-complementarity, i.e. the ability of single strands to form intramolecular base paired structures. Nevertheless, until other techniques such as X-ray diffraction, Raman scattering, or NMR studies on oligomers find evidence for sequence dependence of RNA structure, focus should be on monomer parameters as a source of error.

Possibly the omission of n - \pi^* transitions is a critical factor. However, the calculation for poly AU:AU shown in Figure (17) is a good example of how adjustments in \theta values alone for the present arrangement of \pi - \pi^* transitions can give quite unexpected
CD patterns. In this case, a small maximum at 282 nm and a minimum at 273 nm arise from coupling of monomer transitions located at 260 nm. The measured CD of poly dA:dT is similar in shape and could very likely be correctly predicted by similar alterations in \( \pi - \pi^* \) transition parameters. These calculations have not provided any additional insight into the question of whether inclusion of \( n - \pi^* \) transitions will be necessary or not.

The possibility that the interior environment of the polymer affects the monomer absorption bands, i.e. that static fields are an important influence, should not be ignored. Inability to correct a calculated CD spectrum by varying \( \theta \) values and monopoles generally suggests that shifts in monomer transitions or addition of extra bands should be tested. The alternate resolution of adenine that gives good results for poly A:U and poly AU:AU in RNA geometry is unlikely to be correct, but it illustrates the type of approach that may be needed. Empirical attempts to alter monomer properties should be attempted only when the new parameters can be reliably tested on a number of polymers of reasonably well known geometry. If static fields do prove to be important, dimers may become special cases, since the bases in dimers are more open to solvent.

The attempts to vary monomer parameters that have
been presented here do not represent a comprehensive study of all reasonable parameter changes. They are meant, as an extension of the oligo A studies, to give an indication of how calculated spectra might be expected to change when a given type of parameter change is made. Any more or less complete study of variations in monomer properties would be a rather monumental undertaking. Besides being time consuming, this approach is apt to produce a number of different sets of parameters equally good in reproducing measured spectra. (For example, calculations on heterogeneous dimers, in which \( \theta \) was simultaneously varied for a number of oscillators, showed this to be a problem.) More experimental or theoretical studies on monomer properties are badly needed, as is corroboration by other techniques of proposed solution geometries.

Based on the limited success of the RNA calculations, very tentative speculations can be made concerning the possible sequence dependence of DNA structure in aqueous solution. No combination of monomer parameters with B-DNA geometry came close to predicting the CD of poly dA:dT. Earlier X-ray diffraction studies found anomalous scattering patterns for poly dA:dT that did not belong to the B-DNA family, but Arnott and coworkers attributed this to the presence of poly dT:dA:dT. If the triple strand were the predominant form in solution, these calculations suggest that the
measured CD would have the same shape as would arise from the double stranded structure alone, but with lower magnitude. On the other hand, disproportionation of poly dA:dT in solution would give equimolar amounts of poly dA and the triple strand polymer. Comparison of the calculated CD for poly dT:dA:dT and the measured CD for poly dA (described in Chapter 3, section III) does suggest that the anomalous measured CD of poly dA:dT might arise from a mixture of species. Experimental evidence concerning the stability of double strand versus triple strand conformations lends very little support to this hypothesis. The possibility remains that poly dA:dT may assume either the B conformation or another undefined geometry, depending on the specific solution conditions. Contamination by some structure other than the triple strand is also a possibility. Gray and Bollum have demonstrated, for instance, that contamination is often a problem in measuring the CD of poly dG:dC.

The proposed coordinates of Arnott et al. for poly dAT:dAT and poly dGC:dGC are very appealing, since B-DNA coordinates have not given good results for these sequences. If the calculations are roughly correct, poly dAT:dAT may adopt a solution geometry somewhere between B-DNA geometry and the X-ray coordinates. Possibly the high salt rather than low salt form of poly dGC:dGC is responsible for the observed
fiber diffraction patterns. B-DNA coordinates give very good results for poly dG:dC, in contradiction of X-ray evidence\textsuperscript{27} that this polymer favors the A form. This is all speculation, of course. Nevertheless, it is encouraging to find that different proposed geometries often give rise to widely divergent calculated spectra. Once a reasonably reliable set of monomer properties is determined, CD calculations should be able to confirm or disprove proposed solution structures quite readily.
CHAPTER 4 REFERENCES


28. Johnson, W.C., Jr. and Tinoco, I., Jr. (1969) Biopolymers 7, 727-749. The calculated CD magnitudes for polynucleotides containing two of the four bases should actually be two times larger than published.


Chapter 5
CHAIN LENGTH DEPENDENCE AND THE
NEAREST NEIGHBOR APPROXIMATION

I. **Objective**

The dependence of the calculated CD spectra on chain length is of interest for two reasons. Since the polarizability theory program can handle only a limited number of oscillators, the calculated chain length dependence provides a qualitative estimate of the error involved in comparing spectra calculated for a relatively small number of bases with spectra measured for polymers of much longer chain length. Second, the importance of non-nearest neighbor contributions to CD spectra can be evaluated.

An empirical method, usually referred to as the nearest neighbor approximation, has been developed as an alternative approach to calculating CD or absorption spectra of polynucleotides.\(^1\),\(^2\) For single strands, the assumption is that any optical property \(S\) of a polymer with \(N\) bases (\(N\) a large number) can be approximated at a given wavelength by

\[
S(N)_{\text{total}} = \sum_{i=1}^{4} \sum_{j=1}^{4} X_{ij} S_{ij} - \sum_{i=1}^{4} X_i S_i \tag{37}
\]

where \(X_{ij}\) and \(X_i\) are, respectively, the number fraction of nearest neighbor base combinations \(ij\) and monomers.
i contained in the polymer. $S_{ij}$ and $S_i$ are, respectively, the optical properties due to the interaction between nearest neighbors $i$ and $j$ and to the base $i$ alone. End effects have been neglected. More precisely, if all optical properties are treated on a "per mole of nucleotide residue" basis, the nearest neighbor approximation says that the CD of oligo $A_N$ and poly A, for instance, should be given by

$$CD_{of\ A_N} = \frac{(N-1)2(CD\ of\ ApA)}{N} - \frac{(N-2)(CD\ of\ A)}{N} \tag{38}$$

$$CD\ of\ poly\ A = 2(CD\ of\ ApA) - (CD\ of\ A) \tag{39}$$

Double strands can be treated in the same manner. If the CD due to the entity $\overrightarrow{ApA}$ is abbreviated $\overrightarrow{AA}$, the CD of poly $A:U$ should be given by

$$CD\ of\ poly\ A:U = 2(CD\ of\ \overrightarrow{AA}) - 1/2(CD\ of\ A + CD\ of\ U) \tag{40}$$

(The arrow in this discussion will always be $5'\rightarrow 3'$ and will imply the presence of complementary bases on the opposite strand.) Nearest neighbor terms, e.g. $\overrightarrow{AA}$, $\overrightarrow{AU}$, $\overrightarrow{AG}$, etc., can be extracted from the measured CD spectra of a suitable set of polymers with known nearest neighbor frequencies. (Only 10 of the 16 double strand nearest neighbors are linearly independent, since, for instance, $\overrightarrow{AG} = \overrightarrow{CU}$ and $\overrightarrow{AA} = \overrightarrow{UU}$.) In turn, these nearest neighbor terms can be used to (1)
determine nearest neighbor frequencies for molecules of unknown sequence from the measured CD spectra, and (2) predict the CD for molecules of known nearest neighbor frequency or of known base composition and random sequence.

The validity of the nearest neighbor approach is based on two assumptions. First, of course, the assumption must be made that non-nearest neighbor interactions are not very important or cancel, and second, the geometry of a given nearest neighbor sequence must be the same in the polymers from which the terms are extracted and in the polymers for which data are being calculated. The nearest neighbor method applied to double strand DNA has been found to work well in predicting the CD or nearest neighbor frequencies of most random sequence DNA's and of certain non-random sequence polynucleotides. The approach has been less successful in predicting double strand RNA CD spectra from nearest neighbor terms taken from measurements on synthetic ribo oligomers. In addition, deoxy entities extracted from measurements on polymers in which purines and pyrimidines alternate or are mixed on each strand do not work well in predicting the CD of poly dA:dT, poly dG:dC, and poly AG:TC. This failure has been explained by hypothesizing that poly (pur):(pyr) sequences adopt a structure different from B-DNA geometry. Chain length calculations for
simple sequence polymers have provided new insight into other possible reasons for the occasional failure of the nearest neighbor approximation.

II. Results

The chain length dependence of oligo A\textsubscript{N} CD calculations has already been shown, in figure (5) of Chapter 3. Calculations for varying chain lengths of poly A:U, poly G:C, poly AU:AU, and poly GC:GC in RNA geometry are shown in Figure (19). Chain lengths of 2, 3, 6, and 10 base pairs were calculated for the non-alternating sequences, and chain lengths of 2, 4, 6, and 10 base pairs were done for the alternating sequence polymers. The nearest neighbor composition changes as chain length is increased for the alternating sequences but not for the non-alternating sequences. Most evident is the significant change in \( \lambda_{\text{max}} \) of the longest wavelength band that occurs between chain lengths of 2 and 3 in the non-alternating sequences and in oligo A. For N>3, the differences in calculated CD are mainly in magnitude. The difference between the calculated CD spectra for 2 and 4 base pairs in poly AU:AU is largely in magnitude, but for poly GC:GC, a change in band pattern below 260 nm is produced. For all of these sequences as well as for oligo A, increments in chain length for lengths greater than 6 base pairs involve on the whole only small
Figure 19. Chain length dependence for the calculated CD spectra of (a) poly A:U, (b) poly AU:AU, (c) poly G:C, and (d) poly GC:GC in RNA geometry; the spectra are labeled by numbers indicating the number of base pairs used in the calculation.
(a) Poly A:U

\[ \epsilon_L - \epsilon_R \]

Wavelength (nm)

220
240
260
280

0
10
20
30

2
3
6
10
(c) Poly G:C
changes in magnitude.

Chain length calculations were also done for these sequences in B-DNA geometry but are not shown. For poly A:U, poly G:C, and poly GC:GC, the CD changes from 2 to 3 to 10 base pairs were very similar to, although somewhat smaller than, the changes observed in the corresponding RNA calculations. For poly AU:AU in B-DNA geometry, virtually no changes were observed in values of $\lambda_{\text{max}}$, only changes in magnitude.

Since the intrinsic CD of the monomers is assumed to be very small, eqs. (39) and (40) predict that the CD of poly A and of poly A:U should approach $2(CD$ of ApA) and $2(CD$ of $\overline{AA})$, respectively, as the chain length approaches infinity. That is, the ratio $(CD$ for chain length $N)/(CD$ for chain length 2) should approach 2 as $N \to \infty$. Using the maximum value of $\epsilon_L - \epsilon_R$ at the first long wavelength peak as an indicator, ignoring shape changes, this ratio was calculated for both polymers. Results are shown in Figure (20). It is apparent that this ratio will not reach a value of 2 for the calculated CD even for very long chain lengths.

The double strand calculations for 2 base pairs are theoretically comparable to the double strand nearest neighbor entities, $\overline{AA}$ etc., extracted empirically from measured spectra. Comparison of the calculated CD for the 10 linearly independent combinations of 2 base pairs, in RNA geometry, with the nearest
Figure 20. The calculated ratio \( \frac{\text{CD for chain length } N}{\text{CD for chain length } 2} \) using values for \((\varepsilon_L - \varepsilon_R)_{\text{max}}\) at the first long wavelength peak. This ratio is shown for poly A:U of N base pairs compared to AA, and for oligo A of N bases compared to ApA, in both RNA and B-DNA geometries. Shape changes are ignored.
neighbor terms determined from RNA oligomers\(^6,^9\) showed very poor agreement. Calculated and measured values for \(\text{AA}\) were roughly the same above 220 nm, and \(\text{AU}\) and \(\text{GC}\) showed some similarity at long wavelength. Other combinations were in total disagreement. The 10 combinations calculated in B-DNA geometry were also compared to the ten nearest neighbor contributions determined from DNA CD measurements.\(^3\) Agreement was partially good for \(\text{AG, GG, GC, and TA}\), and poor or non-existent for all other pairs.

III. Discussion

The limitation in maximum chain length that can be calculated by the polarizability theory program does not appear to pose a problem for calculating CD spectra of simple sequence polymers. Calculations for longer chain lengths would likely produce spectra of slightly different magnitudes but virtually the same shape as 10 base pair double strand or 20 base single strand calculations.

On the other hand, these calculations show that non-nearest neighbor interactions can be significant in polymers of simple sequence. Especially when triplets of identical bases appear in a given sequence, exciton-type coupling between identical transitions on these bases may give a substantially different CD pattern than would arise from only two adjacent iden-
tical bases. Triplets of identical base pairs, of the type $\text{AUA}$, will also give somewhat different CD patterns than the sum of $\text{AU}$ and $\text{UA}$. In contrast, the CD of sequences in which $\text{A-U}$ or $\text{G-C}$ base pairs alternate or appear together only in doublets should be approximated much better by the nearest neighbor method. The apparent reason for this is that the strength of coupling between two oscillators is determined by their frequency separation as well as the separation in space of the bases on which they occur. $\text{A}_{260}$ and $\text{U}_{262}$ oscillators can couple with identical oscillators on next nearest neighbor bases, either on the same strand or on the opposite strand. Similarly, $\text{G}_{281}$ can couple with other $\text{G}_{281}$ oscillators at next nearest neighbor positions. However, the coupling of $\text{A}_{260}$ oscillators with $\text{G}_{281}$ oscillators on non-adjacent bases will be much weaker.

Matrix method calculations are useful for checking this hypothesis, since $C_{ik}$ coefficients of monomer band contributions to each polymer $R_K$ band are explicitly calculated. Comparison of matrix method calculations for double strand trimers and dimers in RNA geometry showed, for instance, that the $R_K$ pattern of $\text{AUC}$ was approximately the sum of the $R_K$ patterns calculated for $\text{AU}$ and $\text{UC}$. On the other hand, $\text{AAA}$ was much different from $\text{AA}$, and $\text{AUA}$ was not the sum of $\text{AU}$ and $\text{UA}$. A matrix method calculation of the sequence
TGCTACCGAA showed the CD to result from a large number of small $R_K$ bands, each arising largely from a pair of nearest neighbor base pairs, except for the run of 3 G-C base pairs.

It is evident that nearest neighbor CD patterns extracted from polymers of non-random sequence, e.g. AA from poly dA:dT, will likely prove inadequate in predicting the CD of random sequence polymers. Conversely, nearest neighbor terms extracted from random sequence polymers should not be expected to accurately predict the CD of poly A:T. The fact that the poly A:T CD spectrum has been very poorly reproduced by the nearest neighbor approach may, in fact, be due to unusual conformations for poly (pur):(pyr) sequences in solution. It may equally well be due to the exceptional importance of non-nearest neighbor interactions in these sequences. It is conceivable that the nearest neighbor method would work reasonably well for any sequence if, for instance, two additional terms, namely the AAA and GGG sequences, were added to the current group of linearly independent nearest neighbor terms.

Disagreements between empirically determined nearest neighbor CD terms and CD calculations for 2 base pair sequences do not at this point prove either technique to be wrong. Only a very limited number of polynucleotides of known nearest neighbor frequency
are available for use in determining empirical terms. For the same reason, the nearest neighbor approximation has been tested on only a very few polymers that were not used to determine the nearest neighbor contributions in the first place.\textsuperscript{3-6} Moreover, the ribo terms were determined from oligomers of very high AA content. In either the ribo or deoxy studies, errors in determining a few of the contributions could create large errors in determining the rest. Without a large number of sequences available for testing the parameters, the nearest neighbor terms published so far should be considered as unreliable as the calculated terms.
CHAPTER 5 REFERENCES


CHAPTER 6

POLY I AND POLY G CALCULATIONS:

SINGLE AND MULTI-STRANDS

I. Objective

Guanine is unique among the four common bases in that it is capable of base pairing with other guanine molecules at neutral pH. Under appropriate conditions, guanylic acid monomers in aqueous solution form viscous gels. Figure (21) illustrates two possible ways in which base pairing might occur. X-ray studies on gels of the 5' and 3' isomers of guanylic acid\(^1,2\) and of guanine nucleosides\(^3\) indicate that stacks of monomers form helical structures with the type of base pairing shown in Figure (21a).

Poly guanylic acid (poly G) and poly inosinic acid (poly I) have also been studied by a variety of physical techniques, reviewed in refs. (4) and (5). Inosine differs from guanosine only in the loss of the NH\(_2\) group at the C2 position. The overall evidence suggests that in low salt solution these polymers are in an ordered single strand conformation of unknown form. At higher ionic strength, a stable multi-strand structure appears, probably composed of four parallel strands in similar conformation to the helical arrangement of guanylic acid monomers in gels. Base pairing in the case
Figure 21. Hypothetical base pairing schemes for guanine in guanylic acid gels or between strands of poly G: (a) 4 mononucleotides, or 4 strands of poly G; (b) 3 mononucleotides, or 3 strands of poly G.
of poly I is thought to be the same as in poly G but with one less base pair per base, i.e. with only the O6 and N1 atoms hydrogen bonded.

CD measurements on poly I in low salt and in high salt show two distinct patterns that are dependent on ionic strength, temperature, pH, and time. The high salt CD spectrum is characterized by a large negative band at about 280 nm followed by a positive band near 253 nm. On the supposition that exciton coupling of identical transitions arranged in a helical right handed array will give a positive then negative CD pattern, Thiele and Guschlbauer have proposed that poly I forms a four stranded structure of left handed conformation. However, X-ray studies on poly I support a right handed structure of 3 or 4 strands, most likely of 4 strands. Calculations were therefore done for both poly I and poly G using the X-ray coordinates for poly I in order to determine whether the measured high salt CD pattern could arise from right handed helices. This was the first time such calculations have been attempted.

II. Monomer Properties

The absorption spectrum for inosine at pH 6 is shown in Figure (22). Like the absorption spectrum of guanylic acid (Figure (2)), it is clearly
Figure 22. Resolution II of the measured$^{11}$ inosine absorption spectrum, pH 6.0.
\[ \varepsilon \times 10^{-3} \text{ mole}^{-1} \text{ cm}^{-1} \]
composed of two absorption bands above 230 nm, but the relative strengths of the two bands are not obvious. Three resolutions, labeled I, II, and III, are listed in Table (5) along with corresponding specifications for the guanylic acid resolution. Resolution II is also shown in Figure (22).

CNDO monopoles calculated specifically for hypoxanthine, the base belonging to the nucleoside inosine, were used for these two transitions. (See Table (6).) In accordance with the assignment of guanine monopoles that gave best results for double strand calculations (see Chapter 4, section VIII), initial calculations were done assigning the hypoxanthine monopole set I of ref. (12) to the longest wavelength oscillator and monopole set II to the higher energy oscillator. Crystal or film measurements of transition angles have not been done for hypoxanthine, so the experimentally measured values for the corresponding guanine transitions were used. Since CNDO calculations predict very similar electronic properties for guanine and hypoxanthine, the use of guanine data when necessary was considered justifiable.

III. Poly I Calculations

Calculations were first done using polarizability theory, with an effective dielectric constant
<table>
<thead>
<tr>
<th></th>
<th>$\lambda_{\text{max}}$</th>
<th>$\varepsilon_{\text{max}}$</th>
<th>$\lambda_{\text{max}}$</th>
<th>$\varepsilon_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inosine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resol. I</td>
<td>273</td>
<td>$2.60 \times 10^3$</td>
<td>249</td>
<td>$12.28 \times 10^3$</td>
</tr>
<tr>
<td>II</td>
<td>266</td>
<td>$3.80 \times 10^3$</td>
<td>246.5</td>
<td>$11.38 \times 10^3$</td>
</tr>
<tr>
<td>III</td>
<td>262</td>
<td>$5.0 \times 10^3$</td>
<td>246</td>
<td>$9.82 \times 10^3$</td>
</tr>
<tr>
<td>Guanylic Acid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>281</td>
<td>$6.5 \times 10^3$</td>
<td>252</td>
<td>$13.6 \times 10^3$</td>
</tr>
</tbody>
</table>

Table 5
Absorption Spectra Resolutions
Table 6

CNDO Monopoles for Inosine

| $\theta^\circ$ | $|\mu_1|$ | N1   | C2   | N3   | C4   | C5   | C6   |
|--------------|----------|------|------|------|------|------|------|
| I            | -141.22  | 2.32 | -0.05600 | 0.03360 | -0.01240 | -0.07400 | 0.064100 | 0.012900 |
| II           | -75.56   | 2.22 | 0.034200 | 0.168700 | -0.142800 | 0.168200 | -0.188200 | -0.010000 |

| (I)          |          |      | 0.174600 | -0.293200 | 0.075900 | 0.074500 |
| (II)         |          |      | 0.054600 | -0.069700 | -0.008700 | -0.006300 |

$\theta$ represents the transition angle as calculated by eq. (36) from these monopoles ($q$).

$\mu_1$ values do not represent any resolution of inosine absorption.
of 2, and the 4 strand coordinates of Arnott, Chandrasekaran, and Martilla.\textsuperscript{9} Bases in this coordinate system are related by an axial rise of 3.41 Å and an angle of 31.3°, and are characterized by a negative tilt of -8.9°. Since the current capacity of PROGRAM ROTOPM is 140 oscillators, 4 strands with 11 bases per strand (about one turn of the helix) could not be accommodated if 7 oscillators per base were used. Calculations were therefore done with only two oscillators per base, i.e. using only the two 260 nm region oscillators. Results using inosine resolution II are shown in Figure (23). For comparison, the CD of poly I measured at 3°C in 1.0 M NaCl, after gradual addition of NaCl,\textsuperscript{5} is also shown. This was considered representative of the CD spectra measured for poly I in high salt, but published spectra all differ slightly depending on exact solution conditions and method of preparation. Agreement is remarkably good above 240 nm. Because only two oscillators were used in the calculation and the total CD must sum to zero, the calculated negative band at 233 nm could not be avoided.

To see whether better agreement could be obtained, calculations using resolutions I and III were also done. For each resolution, the transition angle $\theta$ for each of the 2 oscillators was varied over a 180° range. In addition, the assignment of
Figure 23. Calculated CD spectrum of poly I in the 4 strand coordinates of Arnott et al.⁹ (using inosine resolution II) ———. Measured⁵ CD spectrum for poly I at 3°C in 1.0 M NaCl, after gradual addition of NaCl ———.
transition monopoles, i.e. the use of set I versus set II for the longest wavelength oscillator, was simultaneously tested. The assignment of monopole set II to the lowest energy transition and set I to the next transition (as predicted by the CNDO method) never gave reasonable results using inosine resolutions and any combinations of θ values. Transition angles near the values measured experimentally for guanine, i.e. 176° for the first oscillator and -75° for the second, were found to give best agreement with the measured CD. Figure (24) shows representative spectra illustrating the effect of small variations in θ. In general, magnitudes could often be improved at the loss of agreement in band position. Resolution III often gave good results in terms of shape and magnitude but tended to produce peaks that were too far to the blue. Resolution I gave only a very small negative band, centered at 280 nm, and often produced a shoulder on the red side of the 255 nm band.

IV. Poly G Calculations

A calculation using the same 4 stranded poly I coordinates but guanine rather than hypoxanthine parameters for the two oscillators is shown in Figure (25). The calculated spectrum has features similar to the poly I calculations but with wavelength
Figure 24. Calculated CD spectra for 4-strand poly I using Arnott et al.\textsuperscript{9} coordinates, inosine resolution II, and the following values of $\theta$ for the 266 nm and 246.5 nm oscillators, respectively: (a) 165$^{\circ}$, -70$^{\circ}$; (b) 165$^{\circ}$, -75$^{\circ}$; (c) 165$^{\circ}$, -80$^{\circ}$; and (d) 160$^{\circ}$, -85$^{\circ}$.
Figure 25. Calculated CD spectrum of poly G in the 4 strand coordinates of Arnott et al. using 11 bases per strand with 2 oscillators per base ———–;
3 bases per strand with 2 oscillators per base ——–;
3 bases per strand with 7 oscillators per base
(200 nm region and background oscillators of guanine added) ······.
maxima at longer wavelengths. Since absorption bands for guanylic acid are located further to the red than those of inosine, this is expected.

The effect of ignoring far UV and background transitions was estimated by calculating the CD for 4 strand poly G with only 3 bases per strand. With a total of only 12 bases in the polymer, all 7 guanine transitions could be used. Calculations for 12 base poly G are shown in Figure (25), using both 2 and 7 oscillators per base. The addition of distant oscillators increased the magnitude of the first peak and trough, eliminated the second trough, but did not significantly shift the bands. Similar effects could certainly be expected by adding extra oscillators to the 11 base per strand poly I calculations.

The assignment of monopole sets I and II, along with simultaneous changes in $\theta$ values, was also tested for poly G calculations. When set II was assigned to the 281 nm oscillator and set I to the 252 nm oscillator, only one set of $\theta$ values gave results similar to the poly I calculations. Using $\theta=0^\circ$ for the 281 nm oscillator and $\theta=-120^\circ$ for the other oscillator gave a spectrum with a minimum of -4.3 at 276 nm and a maximum of 8.1 at 254 nm. However, a small positive peak of $\epsilon_L-\epsilon_R=1.0$ at 290 nm could not be eliminated.
V. Poly I in Other Geometries

Arnott et al. proposed that the tilt of the bases found by their X-ray studies of poly I may be responsible for the strong negative CD band at 280 nm. Calculations were done in which the bases were made perpendicular to the helix axis by setting all z coordinates (for atoms positions, u_i, and R_i) equal to zero. Since z coordinates for the atoms are all small, from 0.0 to 0.66 Å, this was not considered a serious distortion. Using inosine resolution II and the standard assignment of oscillator parameters, the calculated spectrum for 4 strand poly I was indeed missing the long wavelength negative band. The calculated CD was essentially zero to 265 nm. The 250 nm peak was virtually unchanged, and the loss of the 275 nm negative band was accounted for by an increase in the negative band at 233 nm. The CD spectrum for poly G calculated in the same coordinates with all z values equal to zero still had a negative band at 280 nm, but the magnitude of this band was greatly reduced. In this case, the positive maximum also shifted to the red by 6 nm and increased in magnitude.

Rich published the first X-ray work on poly I in 1958, concluding that the multi-strand structure was composed of 3 strands arranged as in Figure (2lb). Coordinates were estimated for this structure from
the figure in ref. (9) illustrating the 3 and 4 strand models. Axes drawn through the center of the 4 strand figure such that N9 of hypoxanthine was on the x axis. The axes were then scaled to give coordinates for N9 corresponding to the 4 strand coordinates, i.e. X = 6.55 Å, Y = 0.0 Å, and Z = 0.0 Å. The same axes were then used to determine coordinates for N9 of hypoxanthine in the 3 strand figure, X = 6.014 Å, Y = 0.0 Å, and Z = 0.0 Å. All other hypoxanthine atom coordinates for the three strand conformation were simply found by subtracting (6.55-6.014), i.e. 0.536, from the X coordinate values for the hypoxanthine atoms in the 4 strand conformation. The other two strands were generated by adding 120° and 240° to coordinates for the first strand. Helical parameter of θₕ = 41.6° and an axial rise of 3.4 Å were used.

Three stranded poly I calculated with these coordinates, i.e. with the bases tilted as in the Arnott et al. coordinates but with the helical parameters of Rich, gave a CD spectrum very similar to the 4 strand calculation but with a positive peak and short wavelength negative peak of larger magnitudes (e.g. ε̃ₐ - εₐᵣ = +11.2 at 251 nm). Making the bases planar had the same effect as in the 4 strand calculation. Using tilted base coordinates and the helical parameters of Arnott et al. gave a calculated
CD pattern almost identical to the 4 strand calculation. The first negative band was exactly the same and the other two bands were slightly reduced in magnitude.

Recently, new X-ray studies for poly I were published\textsuperscript{10} in which another 4 strand structure is proposed. Base pairing in this model is the same as in the Arnott et al. model, but coordinates are slightly different and the bases are approximately perpendicular to the helix axis. The calculated CD for 4 strand poly I using these coordinates had a positive band of about $\varepsilon_L - \varepsilon_R = +7.6$ at 251 nm and a negative band at $-7.6$ at 233 nm. The negative band at 272 nm was extremely small, with a maximum value of about $\varepsilon_L - \varepsilon_R = -0.10$.

VI. What Causes the Long Wavelength Negative CD Band?

Poly I and poly dI are the only known homopolynucleotides of long chain length that have CD spectra characterized by a negative band at long wavelength. However, some dimers, e.g. GpG in both ribo and deoxy forms, also have CD patterns of this type. Calculations were done both by polarizability theory and the matrix method to look for the origin of this pattern.

The polarizability theory calculated spectra and the measured spectra\textsuperscript{13,14} for the CD of GpG,
both ribo and deoxy, are shown in Figure (26). Both
calculations have small positive bands at long wave-
length. However, some B-DNA calculations done with
other parameters gave mostly negative CD in that
region. The matrix method calculation for GpG
in B-DNA geometry, using only the 281 and 252 nm
oscillators, is shown in Figure (27). As might be
expected, exciton coupling between identical transi-
tions on the two guanine bases gives 2 pairs of
rotational strength bands $R_k$, both of a positive
then negative pattern as observed in the ApA cal-
culations (see Chapter 3). Because both of the two
near UV transitions are strong and because they are
fairly close together in frequency, coupling between
the 252 nm oscillator on one base and the 281 nm
oscillator on the other base, for instance, leads
to mixing of states and hence to unequal strength
of the two $R_k$ bands in each near UV pair. The
splitting of the 281 nm pair is small enough that
the slightly larger gaussian for the negative band
overshadows the slightly smaller gaussian of the
positive band. Presumably, a similar mechanism is
responsible for the calculation of mostly negative
CD above 270 nm in some polarizability theory calcula-
tions.

GpG calculations done on PROGRAM ROTOPR showed
that changes in $\theta$ values could alter the relative
strengths of the $R_k$ bands in each pair and thereby
Figure 26. (a) CD spectrum of ribo GpG: calculated in RNA geometry -----; measured$^{13}$ at 26°C in 0.01 M Na-Po$_4$ buffer, 0.1 M NaClO$_4$, pH 7.2 --- ---; (b) CD spectrum of deoxy GpG: calculated in B-DNA geometry -----; measured$^{14}$ at 26°C in 0.01 M Na-PO$_4$ buffer, 0.1 M NaClO$_4$, pH 7.2 --- ---.
(b) GpG (B-DNA)
Figure 27. CD spectrum of GpG in B-DNA geometry calculated by the matrix method using band halfwidths of 1.5 kK. \( R_k \) patterns are shown by rectangles (positive bands solid, negative bands open) centered at \( V_k \). Only the 281 and 252 nm oscillators of guanine were used in this calculation. \( R_k \) values are per molecule rather than per nucleotide.
determine whether long wavelength CD would be positive or negative. Different assignments of monopoles had the same effect. Matrix method calculations using the same changes in \( \theta \) and/or monopoles showed that some sets of parameters gave sufficient anisotropy in the 281 nm \( R_K \) pair that either the positive or negative Gaussian could completely hide the other, resulting in a totally positive or totally negative CD near 280 nm. In no case, however, was the positive then negative pattern for each \( R_K \) pair ever upset.

This particular mechanism is apparently not responsible for the CD pattern of poly I. Polarizability theory calculations were done for poly I with the 4 strand coordinates but using only one strand, 2 adjacent strands, 2 strands related by 180\(^\circ\) (opposite strands), or 3 strands. The 1 strand and the 2 adjacent strand calculations gave positive CD at long wavelength. For poly G calculated in the same manner, only the 1 strand calculation gave a positive CD pattern. It became apparent that more than one strand was necessary to obtain negative CD at long wavelength, but the necessary number and arrangement of strands was sensitive to oscillator parameters. Nevertheless, the interaction of 2 monomer transitions closely spaced in frequency was not necessary as it was in the calculation of negative CD for GpG. Using the helical parameters for poly I but replacing hypoxanthine with ade-
nine, a calculation of 4 stranded poly A using only
the 260 nm transition of adenine gave a CD pattern
with a large negative band at 255 nm and a positive
band of equal intensity at 242 nm;

Matrix method calculations showed that the $R_K$
pattern arising from the A$_{260}$ transition in 4 stranded
poly A was about the same as the pattern from the G$_{281}$
oscillator in 4 stranded poly G but not the same as
the G$_{252}$ oscillator $R_K$ pattern. The locations of the
$R_K$ bands were about the same regardless of the number
of strands used, but relative strengths of the bands
changed, resulting in different overall CD patterns
for different numbers of strands. In other words, $R_K$
patterns in these multi-strand structures are deter-
mined mainly by the chosen parameters of the monomer
transitions involved, but final CD is dependent on
the number of strands.

VII. Discussion

It is quite clear that the observed CD spectrum
for poly I in high salt can originate from a parallel
arrangement of 4 right hand helices. The calculated
CD is consistent with the proposed 4 strand coordinates
of Arnott et al., but the calculations are not yet
accurate enough to rule out a similar 3 stranded struc-
ture. The calculations do, however, shed some doubt
on the proposal by Zimmerman et al., that the bases
are perpendicular to the helix axis, since a large negative CD band at long wavelength was calculated only when the bases were tilted.

Since calculations of either poly I or poly G using only 1 of the 4 strands gave a large positive band at long wavelength, it seems likely that the helical parameters of the single strands in low salt are different than the parameters of the base paired strands at high salt. Perhaps this is a factor in the complex annealing process observed experimentally as salt concentration is increased.

Agreement between 4 strand poly I calculations and the measured CD spectrum could probably be further improved by placing the first oscillator at slightly longer wavelength. For example, the position of the negative band calculated with resolution I is correct, but the combination of low absorption intensity and weak coupling for this oscillator gives a polymer band that is too small. Addition of 200 nm region and background oscillators might also improve agreement. It is, however, rather difficult to critically judge the calculations when the measured CD spectra vary so much with solution conditions. Very likely the measured spectra represent a mixture of species in solution, so exact agreement can not be expected.

Regardless of accuracy, these calculations prove that CD spectra in which a negative band appears first
at long wavelength can not be taken as indication of any particular geometry. In the case of poly I, the calculated CD pattern is apparently a complicated function of exciton-type interactions between identical oscillators on two or more parallel strands in symmetrical arrays. Many other arrangements might coincidentally give similar patterns. For the double strand 8 fold helix calculation of poly GC:GC (Chapter 4, section VII), a negative CD band appeared first, whereas the same calculation with a 10 fold helix gave a positive band at long wavelength. A proposal that poly IC:IC might be left handed, based on its unusual CD spectrum, is based on insufficient evidence according to these calculations. Particularly when the monopole rather than point dipole approximation is used for calculating $G_{ij}$ values, it is extremely difficult to guess the type of CD pattern that will result for a simple repeating sequence. Predictions are virtually impossible to make when 2 or more non-identical monomer transitions strongly interact. Although all ordered single strand homopolymer calculations carried out so far have produced positive then negative CD patterns, the calculations for GpG suggest that a different choice of monomer parameters might very well produce a negative then positive CD pattern for oligo G calculations.
CHAPTER 6 REFERENCES


Chapter 7

MISCELLANEOUS RESULTS AND SUGGESTIONS

I. Random Sequence Calculations

The fact that the polarizability theory program can handle only a limited number of oscillators makes the calculation of optical properties for truly random sequences impossible with the current program. An attempt was made nevertheless to approximate a random sequence by calculating the CD for a 10 base pair sequence that includes 9 of the 10 linearly independent nearest neighbor pairs. Most calculations were done with the sequence 5'-TGCTACCGA-3' (with complementary bases on the opposite strand). Results are shown in Figure (28) for the CD calculated in RNA, B-DNA, and C-DNA geometries. Parallel and perpendicular components are also shown.

The results for RNA and B-DNA geometries do not agree well with the measured isotropic CD spectra of naturally occurring RNA and DNA. Typical measured spectra can be found in refs. (1-2) and Figure (12). The first peak of the measured CD for RNA has a maximum about 10 nm further to the blue than the calculated spectrum and about half the calculated magnitude. The first peak position is calculated correctly for B-DNA, but the magnitude should be much smaller and the spectrum above 220 nm should be conservative. For these two
Figure 28. Calculated CD spectra for the 10 base pair sequence 5'TGCTACCGAA-3': (a) in RNA geometry, (b) in B-DNA geometry, and (c) in C-DNA geometry. Spectra are shown for the isotropic CD ———; the CD for light incident parallel to the helix axis - - - -; and the CD for light incident perpendicular to the helix axis ······. Isotropic CD is equal to 1/3 (parallel CD + (2 x perpendicular CD)).
(c) Random Sequence (C-DNA)
calculations, first order perturbation theory, which is capable of calculating spectra for much longer chain lengths, give better results in terms of relative magnitudes and shapes, but not in absolute magnitudes or band positions.\textsuperscript{1,2}

In the RNA calculations, the parallel component is non-conservative, as measured experimentally\textsuperscript{3} (Figure (12)), but is too far to the red. Presumably a correct calculation of band position for the parallel component would correct the isotropic CD calculation as well. In the B-DNA calculations, the perpendicular component is much too large compared to measurement\textsuperscript{3} and is apparently responsible for the lack of negative calculated CD near 240 nm. The parallel component is in reasonably good agreement with experimental results.

The calculated CD for DNA in C-DNA geometry does resemble measurements for DNA in ethylene glycol solution\textsuperscript{4} and in LiCl films at low humidity,\textsuperscript{5} although the large negative band is not correctly positioned. On the other hand, neither the perpendicular nor parallel component bears any resemblance to components measured in 6 M LiCl or ethylene glycol.\textsuperscript{3}

There is no way of determining for certain whether or not the limited chain length is at fault in these calculations. However, calculations in which CD spectra were calculated for each of the 32 independent double strand trimers then added and averaged gave
essentially the same results for RNA and B-DNA geometries. (Some minor computational errors were present in the B-DNA trimer calculations, but corrected results would probably not differ significantly.) Calculations were also done for two other 10 base pair sequences, ATTCCGAGGT and TTAGTCTTGG, both of which gave similar results. In addition, calculations were done for the sequences TGCTACCGAA, GCTATCAGGTTGCTACCGAA, and TGCGGATTACTGCTACCGAA, omitting the background oscillators. All three sequences gave nearly identical results. Some negative CD is gained in the B-DNA calculations by dropping background oscillators, but not enough. Based on this evidence, it seems equally likely that disagreement between calculated and measured spectra is caused by errors in monomer property assignments as by the fact that only short sequences can be calculated. Final judgement, however, must wait until better agreement can be obtained in calculations of non-random sequences.

II. Other Double Strand Sequences

CD calculations were also done for a few non-random sequences, other than those discussed in Chapter 4, for which measured spectra are available. Results were mixed. Calculations for 10 base pair poly GT:AC and poly TC:GA gave fairly good agreement with measured CD spectra of the corresponding deoxy poly-
nucleotides. Magnitudes were about 2 times too large at some wavelengths, and band shapes were not correct at other wavelengths. The calculation for poly ATC:GAT (9 base pairs) predicted an approximately correct CD shape, but magnitudes 2 to 3 times too large. Nine base pair calculations of poly TAC:GTA and poly AAC:GTT were not very successful. In the latter case, the CD at long wavelength had features that resembled the measured CD of poly dA:dT, whereas the calculated CD was much like the calculated CD of poly A:U in B-DNA geometry. A calculation done for 10 base pair poly AC:GU in RNA geometry was correct at some wavelengths but had the wrong sign near 270 nm. However, this particular calculation has only been done with $\varepsilon_{\text{dielectric}} = 1$ and might improve considerably with use of a higher dielectric constant.

III. Dimer Calculations

Calculations have already been presented for ApA (Chapter 3) and GpG (Chapter 6) in both RNA and B-DNA geometries. Calculations for the other 14 dimers, in both geometries, are shown in Figure (29). Agreement with measured CD$^6,7$ is good in a few cases, shows only some resemblance to measured spectra in other sequences, and is totally wrong for several others. Since information on dimer conformations in solution is presently very limited, the accuracy of these calculations is
Figure 29. (a-g) Dimer calculations in RNA geometry; (h-n) Dimer calculations in B-DNA geometry. All were done with polarizability theory and $\varepsilon_{\text{diel}} = 2$. 
difficult to judge.

While dimer calculations are of little use right now, they may eventually provide the easiest means of empirically adjusting monomer parameters. Dimers are more easily studied by other techniques such as NMR than are longer polynucleotides. As information is compiled on dimer conformations, calculations can be done for the proposed geometries. In comparison to CD spectra for longer sequences, the calculated CD patterns for dimers are exceedingly uncomplicated. Necessary adjustments in monomer parameters are much more easily ascertained. The discussion of parameter adjustments for ApA calculations, presented in Chapter 3, gives a good example of how problems with monomer parameter assignments can be approached using dimer calculations.

IV. General Suggestions

No convincing evidence has accumulated in these calculations that any one component of the calculations is largely responsible for remaining discrepancies between calculated and measured optical properties. Problems that might be caused by the omission of n - \( \pi^* \) transitions or by ignoring static field and solvent effects are still open to question. (Ref. 9, for instance, applies solvent effects to polypeptide CD calculation.) A different approach to the use of an
effective dielectric constant might give substantially different results for many calculations. It would also not be particularly surprising to find that assumptions concerning geometries in solution have been in error. Nevertheless, monomer properties for in-plane transitions should probably still be considered the most likely major source of error.

Most suspect are the resolutions of monomer absorption spectra and the choice of monopoles. Many alternative assignments of monomer bands are possible that were not explored. For instance, moving the second cytosine transition from 230 nm to 240 nm could have a profound effect on many calculations. Additional bands, particularly in the 200 nm region, could probably be justified and might prove important. The evidence for an additional band in the 240 nm region for uracil, for instance, should perhaps be explored more completely. Also, alternative choices of monopoles were not fully tested in the double strand calculations. Particularly in G/C calculations, some work done with BASES calculated monopoles rather than CNDO monopoles indicated that the BASES generated monopoles might work better for some near UV transitions.

If further attempts are made to adjust the position and number of monomer absorption bands, the matrix method may prove much less cumbersome to use. Shifts in band positions can be accomplished merely by
changing $\tilde{\nu}_i$ values. The matrix method is probably adequate for such a purpose, as long as calculations are done periodically by the polarizability theory program in order to evaluate differences in the CD calculated by the two theories.
CHAPTER 7 REFERENCES

1. Johnson, W.C., Jr. and Tinoco, I., Jr. (1969) Biopolymers 7, 727-749. The calculated CD magnitudes for polynucleotides containing all of the four bases should actually be four times larger than published.


Appendix

COMPUTER PROGRAMS

I. PROGRAM ROTOPM. This is a revised version of the program ROTOP by Werner Hug. ROTOPM (a) calculates $G_{ij}$ by the point monopoles approximation, and (b) is restricted to calculating helical structures. Calculation of all coordinates is by rotation and translation of "reference" coordinates at the base of a helix.

Current Limitations:
- 4 "reference" monomers
- 7 oscillators per "reference" monomer
- 140 oscillators (total)
- 22 point monopoles per oscillator
- 26 sets of polarizabilities
- 101 frequency points

Deck set-up:
1. Control cards
2. Program deck
3. Input deck
   a. geometry specification deck
   b. polarizability deck
   c. multiple polymer specifications decks
4. End of job card

Out of plane monopoles:
Currently placed .75Å above and below plane of base. Alternate spacing may be used by changing card in INPUT subroutine which reads DP=.75

Geometry Specification Deck:

I. THETA, D, B (2F10.3,5A10)  

THETA=angle (deg) between adjacent monomers  
D=distance along helix axis between adjacent monomers  
B= description of geometry for print-out
II. Set of cards for each reference monomer

A. ID,NATOMS,NOSC,NIN
   ID=A,T,G, or C, etc.
   NATOMS=# of atoms on monomer
   NOSC=# of oscillators on monomer
   NIN=# of oscillators for which monopoles are in-plane

   (A1,9X,2(I2,8X),I2)

B. TP,PP
   TP=θ
   PP=ϕ
   (2F10.3)
   Spherical coordinates of unit vector perpendicular to plane of monomer

C. (AR,AT,AZ)
   (9F8.3) - 3 atoms per card
   AR=R AT=θ AZ=Z
   Cylindrical coordinates for atoms on monomer

D. R,T,Z,TE,PE,TM,PM,BI,DB,IPOL
   (9F8.3,6X,I2)
   R,T,Z=R,θ,Z (in cylindrical coordinates)
   of oscillator position on monomer
   TE,PE,TM,PM=θ,ϕ (in sph. coord.) of electric and magnetic oscillator unit vectors, respectively
   BI=βi for magnetic oscillators (=0 if osc doesn't exist)
   DB=|û| in Debye
   IPOL=identification # of polarizability set corresponding to oscillator

E. Q's
   1 set of NATOMS numbers for each NIN osc
   1 set of 2xNATOMS numbers for each NOSC-NIN oscs
   (8F10.6)
   Note: order must agree with atom in C. For out-of-plane q's, give 1st above plane then below plane q's for each atom
   D. and E. in series - 1 set for each oscillator on the reference monomer. Start with all NIN oscillators first (see A.)
End of (II.) series is signified by card with punch in column 80 if less than 4 reference monomers are read in. Otherwise a DO loop in program terminates after 4 monomers have been specified.

Polarizability Deck:

I. Parameter Card

\[ \text{TUA}, \text{TUE}, \text{TUD}, \text{DS}, \text{UK}, \text{GK}, \text{K} \] (6E10.0, I2)

1 TUA=starting frequency (in kK)
2 TUE=ending frequency (kK)
3 TUD=frequency increment (kK)
4 DS=dipole strength (in Debye squared)
5 UK=resonance frequency (kK)
6 GK=damping constant (kK)
7 K=identification number of the polarizability set

4-6 are for damped oscillator specifications. LEAVE THEM BLANK if polarizability versus frequency is to be read in directly, and follow the parameter card by:

II. Imaginary polarizability vs \( \tilde{\nu} \) (in \( \text{A}^3 \)) (10E8.0)

III. Real polarizability vs \( \tilde{\nu} \) (in \( \text{A}^3 \)) (10E8.0)

Repeat \{1, 2, 3\} for all sets of polarizabilities. TUA and TUE may be different, but TUD must be the same for all sets. The program limits the calculation to the maximum frequency range common to all sets. TUA-TUE may be either positive or negative.

End the series with the following card:

Column 1 = 1 or 0 If=1, polarizabilities versus frequency are printed in output

Punch in column 80

Polymer Specification Deck

All "reference" information has now been entered. The following sequence of cards may be repeated for any number of runs (one polymer calculation per run.) End the series with a card with punch in column 80.

I. A,XMER A=title describing polymer
XMER = \# of monomers by which final result will be divided to give, e.g., CD per monomer. Is set equal to 1 if not specified.

II. IGMAT, IPNCH, IPLT, UP (up to 6 numbers)

if IGMAT \neq 0, G matrix is printed
if IPNCH \neq 0, program punches output
if IPLT \neq 0, program plots output
UP = a max of 6 frequency points (all optional) at which the A matrix, before and after inversion, and the final tensor, before transformation, are printed.

III. IBASE, NBR, IOSC (1 for each oscillator making up NMR)

IBASE = reference monomer (A, T, G, or C, etc)
NBR = \# of oscillators (max = \# of oscs specified for IBASE monomer in geometry deck) to be used in run
IOSC specifies the NBR specific oscillators to be used in run as numbered in the reference monomer section of geometry deck.

IV. POS (up to 16 numbers)

POS numbers specify the positions on the helix at which the reference monomer IBASE is to be located. Position 1 is the reference monomer itself. Position 2 translates the ref. monomer by THETA, D once, etc. A negative sign indicates the complementary strand.
E.G. 1. 2. 3. gives positions 1,2,3 on one strand
-1. -2. -3. gives the base pair monomers on the opposite strand

Series III. - IV. may be repeated to give a total of 140 oscillators on any number of positions. End series with a card with a punch in column 80. Hence, program should end with 2 punch-in-col-80 cards and End of Job card.
PROGRAM ROTORM (INPUT, OUTPUT, PUNCH, TAPE5=INPUT, TAPE6=OUTPUT,
1TAPE7=PUNCH, TAPE1, TAPE2)
COMMON/MISC/KOS, NROS, IRUN, JRUN, NRUN, IB, U, UL, UU, DU, IMAT, IGMAT, XMER,
1IPNCH, IPLT, KF
NRUN=0
1 CALL INPUT
IRUN=0
CALL FREQU
NRUN=1
GO TO 1
END
SUBROUTINE INPUT
COMMON/IC/SC/NNOS,IRUN, JRUN, NRUN, IB, U, UL, UU, DU, IMAT, IGAM, XMER,
IPNCH, IPLT, KF
COMMON CI/292000
COMMON/ROSC/XR(140), YR(140), ZR(140)
COMMON/EEOS/XE(140), YE(140), ZE(140)
COMMON/EMOS/ XM(140), YM(140), ZM(140)
COMMON/POL/ALT(26), ART(26), UAI(26), UP(6)
COMMON/IDET/ NDA(140), BSUB(140)
COMMON/MPOL/QI(22,74), DB(17,41), IDB(140), IDDB(140), IQI(140)
COMMON/PROSC/ XR(140), YR(140), ZR(140)
COMMON/POL/ALT(26), ART(26), UAI(26), UP(6)
COMMON/IDET/ NDA(140), BSUB(140)
COMMON/MPOL/QI(22,74), DB(17,41), IDB(140), IDDB(140), IQI(140)
DIMENSION AT(7), BS(5)
DIMENSION NATOMS(4), ID(4), NIN(4), IDSC(12), POS(12)
DIMENSION R(7,4), T(7,4), Z(7,4), TE(7,4), PE(7,4), IPOL(7,4)
DIMENSION TM(7,4), PM(7,4), BI(7,4)
DIMENSION TP(41), PP(4)
DIMENSION AR(11,4), AT(11,4), AZ(11,4)
DIMENSION AR(22,4), AT(22,4), AZ(22,4)
DIMENSION XA(22,140), YA(22,140), ZA(22,140)
EQUIVALENCE (XA(1,1), C(36100)), (YA(1,1), C(33000))
EQUIVALENCE (ZA(1,1), C(29900))
IF (NRUN.NE.0) GO TO 390
READ(5,300) THETA, DB
300 FORMAT (FIO,3,F10.3,5A10)
WRITE (6,99)
99 FORMAT (1H20,1H*,1H8,1H*2,1H*, 1H8,1H*3,1H*, 1H8,1H*4,1H*, 1H8,1H*5,1H*, 1H8,1H*6,1H*, 1H8,1H*7,1H*, 1H8,1H*8,1H*, 1H8,1H*9,1H*, 1H8,1H*10,1H*)
WRITE (6,200)
200 FORMAT (1X,*C, CECH ROTOMP (MONOPOLE INTERACTION)*,45X,*COUPLE
1D OSCILLATOR MODEL TO INFINITE ORDER*)
WRITE (6,201) (B(II), I=1,5)
201 FORMAT (1X,5A10,36X,*OF H. DEVDE Program W. HUG*)
WRITE(6,220)
220 FORMAT (1X,*REFERENCE BASE DATA*)
DO 380 I=1,4
READ(5,301) ID(II), NATOMS(II), NOSC, NIN(II), ISEP
301 FORMAT (A1,9X, I2,8X, I2,8X, I2,47X, I1)
IF (ISEP.NE.0) GO TO 381
READ(5,302) TP(II), PP(II)
302 FORMAT (2(FIO,3))
JJJ=NATOMS(II)
READ(5,303) ((AR(J,II), AT(J,II), AZ(J,II)), J=1, JJJ)
303 FORMAT (9F8.3)
RADT=TP(II)*0.0174533
RDP=PP(II)*0.0174533
XP=SIN(RADT)*COS(RADP)
YP=SIN(RADT)*SIN(RADP)
ZP=COS(RADT)
WRITE (6,221) ID(II)
DO 901 J=1, JJJ
C DP SPECIFIES FRACTION ALONG UNIT VECTOR PERPENDICULAR TO PLANE
C OF BASE THAT MONOPOLE POSITIONS ARE MOVED
DP=.75
K=2*J
L=K-1
RAD = AT(J, I) * 0.0174532
X = AR(j, I) * SIN(RAD)
Y = AR(J, I) * SIN(RAD)
XUP = X * (DP*XP)
XDN = X - (DP*XP)
YP = Y * (DP*YP)
YDN = Y - (DP*YP)
RUP = XUP*XUP+YUP*YUP
RDN = XDN*XDN+YDN*YDN
AR(J, I) = SQRT(RDN)
AR(J, I) = SQRT(RUP)
AZ(J, I) = AZ(J, I) - (DP*ZP)
AZ(J, I) = AZ(J, I) + (DP*ZP)
ADN = ABS(XDN)
AUP = ABS(XUP)
IF (ADN.LT.0.0001) GO TO 910
TODN = YDN/XDN
AT(C, I) = AT(ADN)*57.2958
IF (XDN.LT.0.) AT(J, I) = AT(J, I) + 180.
GO TO 911
910 AT(J, I) = 90.
IF (YDN.LT.0.) AT(J, I) = -90.
IF (YDN.EQ.0.) AT(J, I) = 0.
911 CONTINUE
IF (AUP.LT.0.0001) GO TO 912
TUP = YUP/XUP
AT(L, I) = AT(TUP)*57.2958
IF (XUP.LT.0.) AT(L, I) = AT(L, I) + 180.
GO TO 913
912 AT(L, I) = 90.
IF (YUP.LT.0.) AT(L, I) = -90.
IF (YUP.EQ.0.) AT(L, I) = 0.
913 CONTINUE
WRITE (6, 223) J, X, Y, AZ(J, I), AR(J, I), AT(J, I) + L, XUP, YUP, AZ(L, I),
        1AP(L, I), AT(L, I) * K, XDN, YDN, AZ(J, I), AR(K, I), AT(K, I)
223 FORMAT (7X, 12, 4(F5.2, 1X, F6.2, 1X, 2(J12, F5.2, 1X, F5.2, 1X, F5.2, 1X,
        1F5.2, 1X, F6.2, 1X))
901 CONTINUE
JJ = NIN(I)
KJ = 2*JJ
WRITE (6, 222) JJ, JJ, K
222 FORMAT (/5X, *OSC*, 2X, *DB*, 3X, *MONOPOLES IN ORDER FOR ATOMS 1 THRU
1*I, I2, * IN PL TRANS AND FOR POS I THRU *, I2, * OUT OF PL TRANS*)
IF (JJ.EQ.0.) GO TO 240
DO 382 J = 1, JJ
READ (5, 304) R(J, I), T(J, I), Z(J, I), T(J, I), PE(J, I), TM(J, I), PM(J, I),
1BJ(J, I), DB(J, I), IP(J, I)
304 FORMAT (9F8.3, 6X, I2)
READ(5, 305) LQ(J, I), L = 1, JJ
305 FORMAT (8F10.6)
WRITE (6, 224) J, DB(J, I), LQ(J, I), L = 1, JJ
224 FORMAT (/5X, I2, 1X, F5.2, 2X, 1I11, F8.6, 1X)
382 CONTINUE
240 CONTINUE
IF (JJ.EQ.NOSCI) GO TO 241
K = JJ+1
DO 383 J=K,NDSC
READ (5,304) R(J,1),T(J,1),Z(J,1),TE(J,1),PE(J,1),TM(J,1),PM(J,1),
           1BI(J,1),DB(J,1),IPOL(J,1)
READ(5,305) (QL(J,1),L=1,KJ)
WRITE (6,225) J,DB(J,1),(QL,J,1),L=1,KJ)
225 FORMAT (5X,I2,1X,F5.2,1X,F8.6,1X),/,15X,11(F8.6,1X))
383 CONTINUE
241 CONTINUE
380 CONTINUE
381 CONTINUE
DO 370 I=2,4
IF(I.EQ.2) ID(I)=1HX
IF(I.EQ.3) ID(I)=1HY
IF(I.EQ.4) ID(I)=1HZ
NATOMS(I)=NATCMS(I)
NIN(I)=NIN(I)
J=1
KK=NATOMS(I)
DO 371 K=1,KK
AR(K,I)=AR(K,1)
AT(K,1)=AT(K,1)+(I*90.)
AZ(K,1)=AZ(K,1)
371 CONTINUE
DO 372 L=1,2
RL(L,1)=RL(L,1)
TL(L,1)=TL(L,1)+(I*90.)
Z(L,1)=Z(L,1)
TE(L,1)=TE(L,1)
PE(L,1)=PE(L,1)+(I*90.)
BI(L,1)=BI(L,1)
PB(L,1)=PB(L,1)
IPOL(L,1)=IPOL(L,1)
DO 373 K=1,KK
QL(K,L,1)=QL(K,1,1)
373 CONTINUE
372 CONTINUE
370 CONTINUE
CALL ASHAPE
390 CONTINUE
READ(5,100) (AI(I),I=1,T1),XMER,ISEP
100 FORMAT (7A10,2X,F4.0,3X,I1)
IF (XMER.EQ.0) XMER=1.
IF (ISEP.NE.0) STOP
READ(5,102) IGCI,IPNCH,IPLT,(UP(I),I=1,6)
102 FORMAT (311,7X,6E10.0)
WRITE (6,99)
WRITE (6,230) (AI(I),I=1,T1)
230 FORMAT (7A10)
WRITE (6,206)
206 FORMAT (/2X,*BASE*,2X,*POS ON*,7X,*ELECTRIC OSCILLATOR POSITIONS*,
          123X,*DIRECTION UNIT VECTORS*,15X,*MAG OSC*)
WRITE (6,207)
207 FORMAT (8X,*HELIX*)
WRITE (6,208)
208 FORMAT (120X,*X*,8X,*Y*,8X,*Z*,7X,*R*,6X,*THETA*,2X,*OSC*,1X,
          1*POL*,5X,*X*,8X,*Y*,8X,*Z*,5X,*THETA*,5X,*PHI*,4X,*OSC*,2X,*BI*)
IRUN=0
JRUN=0
NOS=0
IB=0
DO 1 IJK=1,20
READ (5,101) IBASE,NBR,(IOSC(I),I=1,12),ISEP
101 FORMAT (12I1,3X,12I1,3X,9X,11)
IF (ISEP.NE.0) GO TO 50
READ (5,103) (POS(I),I=1,12)
103 FORMAT (12F5.0)
LINE1=0
WRITE (6,210)
WRITE (6,209) IBASE
209 FORMAT (1H+,3X,A1)
DO 3 I=1,4
IF (IBASE.EQ.ID(I)) GO TO 4
CONTINUE
WRITE (6,204)
204 FORMAT (1//I4,3X,AL)
STOP
4 CONTINUE
DO 2 J=1,12
IF (POS(J).EQ.0) GO TO 1
STRAND=1.
WC=1HW
IF (POS(J).GT.0) GO TO 15
STRAND=-1.
WC=1HC
210 FORMAT (1I)
POSJ=ABS(POS(J))
IF (LINE1.EQ.1) WRITE (6,210)
WRITE (6,211) WC,POSJ
211 FORMAT (1H+,9X,A1,F3.0)
POSJ=POSJ-1.
LINE2=0
NNos=NOS
DO 5 M=1,NBR
K=IOSC(M)
NOS=NOS+1
TT=(T(K,I)*STRAND+POSJ*THETA)
TRAD=TT*.0174533
XR(NOS)=R(K,I)*COS(TRAD)
YR(NOS)=R(K,I)*SIN(TRAD)
ZR(NOS)=R(K,I)*STRAND+POSJ*DI
TTE=TE(K,I)
IF (POS(J).LT.0) TTE=180.-TE(K,I)
TERAD=TTE*.0174533
PPE=(PE(K,I)*STRAND+POSJ*THETA)
PERAD=PPE*.0174533
XE(NOS)=SIN(TERAD)*COS(PERAD)
YE(NOS)=SIN(TERAD)*SIN(PERAD)
ZE(NOS)=COS(TERAD)
ID(A(NOS))=IPOL(K,I)
IDD(B(NOS))=K
IF (LINE2.EQ.1) WRITE (6,210)
WRITE(6,212) XR(NOS), YR(NOS), ZR(NOS), R(K, I), TT, NOS, IDA(NOS),
1XE(NOS), Y(E(NOS), ZE(NOS), TTE, PPE
212 FORMAT (1H+,15X,3(F7.4,2X1,1X,F5.2,2X,F7.2,2X,F7.2,1X,F3,1X,12,3(2X,F7.4)
12X,F6.2,2X,F7.2)
   IF(BI(K,I),EQ,0) GO TO 6
   IB=1
   TTM=T(K, I)
   IF(POS(J,I).LE.0) TTM=180.-T(K, I)
   TMRAD=TTM*.0174533
   PPM=(PMK(I)*STRANDI+POSJ*THETA)
   PMRAD=PPM*.0174533
   XM(NOS)=SIN(TMRAD)*COS(PMRAD)
   YM(NOS)=SIN(TMRAD)*SIN(PMRAD)
   ZM(NOS)=COS(TMRAD)
   BSUBI(NOS)=BI(K, I)
   WRITE(6,213) NOS,BSUBI(NOS), XM(NOS), YM(NOS), ZM(NOS)
213 FORMAT (1H+,91X,15,F5.2,F10.4)
   GO TO 9
   6 XM(NOS)=0.
   YM(NOS)=0.
   ZM(NOS)=0.
   BSUBI(NOS)=BI(K, I)
   9 CONTINUE
   MM=NIN(I)
   IF(K.GT.MM) GO TO 10
   NMPTS=NATOMS(I)
   IDQ(NOS)=NMPTS
   DO 8 L=1,NMPTS
       ANGLE=(ATIL,L)*STRAND1+POSJ*THETA
       RAD=ANGLE*.0174533
       XL(L,NOS)=AR(L, I)*COS(RAD)
       YL(L,NOS)=AR(L, I)*SIN(RAD)
       ZL(L,NOS)=(AZL,L)*STRAND1+POSJ*D
   8 CONTINUE
   GO TO 11
10 CONTINUE
   NMPTS=2*NATOMS(I)
   IDQ(NOS)=NMPTS
   DO 7 L=1,NMPTS
       ANGLE=(ATIL,L)*STRAND1+POSJ*THETA
       RAD=ANGLE*.0174533
       XL(L,NOS)=AR(L, I)*COS(RAD)
       YL(L,NOS)=AR(L, I)*SIN(RAD)
       ZL(L,NOS)=(AZL,L)*STRAND1+POSJ*D
    7 CONTINUE
11 CONTINUE
5 LINE2=1
   CALL INTRAC
2 LINE1=1
1 CONTINUE
50 JRUN=1
   RETURN
END
SUBROUTINE ASHAPE
COMMON/MISC/NDOS, NNRUN, IRUN, JRUN, IBI, U, UU, DU, IMAT, XMAT, XMEAT, XPDCH, XPLT, XFK
COMMON/AIR/AL(26, 101), AT(26, 101)
COMMON/POL/AT(26, 101)
DIMENSION UF(26), UD(26), IPTS(26)
EQUIVALENCE (TUA, UP(1)), (TUE, UP(2)), (TUD, UP(3)), (OS, UP(4)),
(UK, UP(5)), (UK, UP(6))
DO 25 I=1, 26
UA(I)=0.
UE(I)=0.
UD(I)=0.
25 CONTINUE
DO 1 I=1, 27
READ(5, 101) TUA, TUE, TUD, DS, UK, GK, K, DMOVE, ISEP
100 FORMAT (6E10.3, 12, 8X, F3.0, 6X, I1)
IF (ISEP.NE.0) GO TO 2
IF (K.EQ.0) K=I
UA(I)=TUA
UE(I)=TUE
UD(I)=TUD
RIZ=(TUE-TUA)/TUD
RAB=ABS(RIZ) + 1.01
IAB=RAB
IPTS(I)=IAB
IF (IAB.GT.101) GO TO 5
IF (UK.NE.0) GO TO 3
READ (5, 101) (AI(K, J), J=1, IAB)
101 FORMAT (2(E8.0, 32X))
IF (DMOVE.NE.0) GO TO 7
UA(I)=TUA-(DMOVE*TUD)
TUE=TUE-(DMOVE*TUD)
UE(I)=TUE
GO TO 7
3 IF (GK.EQ.0) GK=0.5*SQRT(UK)
DO 6 J=1, IAB
UDEL=(J-1)*TUD
IF (RIZ.LT.0) UDEL=-UDEL
U=TUA+UDEL
UK=UK+U
UU=UU+U
UMU=UKU-UU
UU=DS*UK*10.069/(UMU*UMU+UU*GK*GK)
TBM=UMU+UU
IF (ABS(TBM).GT.1000000.) GO TO 5
AR(K, J)=TBM
TBM=U*GK*UU
IF (TBM.GT.1000000.) GO TO 5
AR(K, J)=TBM
6 CONTINUE
7 CONTINUE
IF (RIZ.GT.0) GO TO 1
IA=IAB/2
DO 4 J=1, IA
JT=IAB+1-J
4 CONTINUE
TBM = AI(K, J)
AI(K, J) = AI(K, JT)
AI(K, JT) = TBM
TBM = AR(K, J)
AR(K, J) = AR(K, JT)
AR(K, JT) = TBM

4 CONTINUE
UE(K) = UA(K)
UA(K) = TUE
1 CONTINUE
5 WRITE (6, 220)
220 FORMAT (1X, *POL OUT OF RANGE*)
STOP
2 UL = 0.
UU = 10E5
DO 22 I = 1, 26
UL = AMAX1 (UL, UA(I))
UT = AMIN1 (UU, UE(I))
IF (UT .NE. 0.) UU = UT
IF (UD(I) .EQ. 0.) GO TO 22
IF (I .LT. 26) I I = I + 1
DU = UD(I)
IF (UD(I) .EQ. 0.) GO TO 22
IF (UD(I) .EQ. UD(I)) GO TO 22
WRITE (6, 270)
270 FORMAT (1X, *POL INCOMPATIBLE*)
STOP
22 CONTINUE
IF (TUA .NE. 0.) GO TO 30
RETURN
30 CONTINUE
WRITE (6, 200)
200 FORMAT (1X, *POLARIZABILITIES*)
DO 31 I = 1, 26
IF (UD(I) .EQ. 0.) GO TO 31
WRITE (6, 201) I, UA(I), UE(I), UD(I)
31 CONTINUE
IF, T, * DELTA = *, F7.3)
IAB = IP(S(I))
WRITE (6, 204)
204 FORMAT (1X, *IMAGINARY*)
WRITE (6, 205) (AI(I, J), J = 1, IAB)
205 FORMAT (1X, 10F12.5)
WRITE (6, 206)
206 FORMAT (1X, *REAL*)
WRITE (6, 205) (AR(I, J), J = 1, IAB)
31 CONTINUE
RETURN
SUBROUTINE FREOU
COMMON/MISC/NOS, NVOS, IRUN, JRUN, NRUN, IB, UL, UU, DU, IMAT, IGMAT, XMER,
IPNCH, IPT, KF
COMMON/MDAT/IMAT
COMMON/MD1/AR(26,101), AR(26,101)
COMMON/MD2/AR(26,101), AR(101), UP(6)
COMMON/WWW/WWW(101,4)
DIMENSION EXT(101), CD(101), ORD(101), FREQ(101)
EQUIVALENCE (WWW(1,1), EXT(1), WWW(1,3), CD(1)), (WWW(1,4), ORD(1))
EQUIVALENCE (WWW(1,2), FREQ(1))
IMAT=0
NI=1.01 + (UU-UL)/DU
DO 1 J=1,NI
UJ=UL+(J-1)*DU
DO 2 I=1,26
IF (UA(I),EQ,0.) GO TO 2
K=1.01 + (U-UA(I))/DU
KF=K
ARTI=ARTI
ARTI=AR(I,K)
2 CONTINUE
IMAT=0
D2=DU/2.
DO 86 NN=1,6
UTEST=UPCNI-U
IF (UTEST*CUTEST-D2t.LE.O . OR.UTEST*CUTEST+DZt.LT.O.t)
IMAT=1
86 CONTINUE
CALL INTRAC
IF (IRUN.NE.O) GO TO 3
WRITE (6,201) XMER
WRITE (6,202) U
3 CONTINUE
201 FORMAT (///,**OPTICAL DATA**)
202 FORMAT (/s,3X,*FREQU*,9X,*TRACE*,11X,*AXIS*, X, Y, Z*,17X,
1 *AXIS 2, X, Y, Z*,17X,*AXIS 3, X, Y, Z*,17X)
203 FORMAT (1X,*TRACE ALREADY DIVIDED BY NUMBER OF MONOMERS - *,F4.01
WVLTH=1D*41/U
WRITE(6,199) WVLTH
199 FORMAT (1X,F7.3)
WRITE (6,200) U
200 FORMAT (1H,F7.3)
CALL AMAT
CALL UVSHAPE
CALL RTSHAPE
IRUN=IRUN+1
1 CONTINUE
IF (IPNCH.EQ.0) GO TO 20
DO 4 K=1,4
4 WRITE(7,204) (WWW(I,K),I=1,NI)
204 FORMAT (8F10.3)
20 CONTINUE
IF (IPPLT.EQ.0) GO TO 11
DO 12 I=1,NI
EXT(I)=WWW(I,1)
CD(I)=WWW(I,3)
ORD(I)=WW(I,4)
FREQ(I)=UL+(I-1)*DU
12 CONTINUE
ISX=0
IF(INI.LT.101) GO TO 13
ISX=1
NI=100
13 CONTINUE
900 FORMAT (1H1)
WRITE (6,900)
WRITE (6,901)
901 FORMAT (1X,*EXTINCTION COEFFICIENT PER MONOMER*)
CALL PRNPLT(FREQ,EXT,55.,.3,25000.,500.,ISX,0,NI)
WRITE (6,900)
WRITE (6,902)
902 FORMAT (1X,*CD (E LEFT - E RIGHT) PER MONOMER*)
CALL PRNPLT(FREQ,CD,55.,.3,30.,1.,ISX,0,NI)
11 CONTINUE
RETURN
END
SUBROUTINE INTRAC
COMMON C(39200)
COMMON/MISC/NOS,NNOS,IRUN,IRUN,NNRUN,IB,UL,UL,DU,IMAT,IGMAT,XMER,
IPNCH,IPLT,KF
COMMON/MPOL/G(22,7,4),IDB(140),IDDB(140),IDQ(140)
DIMENSION XA(22,140),YA(22,140),ZA(22,140)
EQUIVALENCE (XA(I,1)),C(36100)11,(YA(I,1),C(33000))
EQUIVALENCE (ZA(I,1),C(12990))
DIMENSION R(3)
DIMENSION G(140,140)
EQUIVALENCE (C(I),G(I,1))
IF(IRUN.NE.0) GO TO 10
IMAT=0
IF(NNOS.GT.0) GO TO 26
DO 25 I=1,NOS
DO 25 J=1,NOS
25 G(I,J)=0.
RETURN
26 CONTINUE
NN=NNOS+1
DO 1 I=NN,NOS
II=IDB(I)
KK=IDQ(I)
DO 2 J=1,NNOS
JJ=IDB(J)
LL=IDQ(J)
GG=0.
DO 3 K=1,KK
DO 4 L=1,LL
R1=XAK(I,1)-XAL(J)
R2=YAK(I,1)-YAL(J)
R3=ZAK(I,1)-ZAL(J)
DD=0.
DO 5 M=1,3

5 DD = DD + R(M) * R(M)
   D = SQRT(DD)
   M = IDDB(I1)
   N = IDDB(J1)
   GG = GG + [(Q(K, M, II) * Q(L, N, JJ)) / D]
4 CONTINUE
3 CONTINUE
   G(I, J) = (4.80298 * 4.80298 * GG) / (DB(M, II) * DB(N, JJ))
   G(I, J) = G(I, J) / 2.0
   G(J, I) = G(I, J)
2 CONTINUE
   DO 1 J = NN, NOS
   G(I, J) = 0.
   G(J, I) = 0.
   1 CONTINUE
   RETURN
10 CONTINUE
   REWIND 1
   IF (IRUN .NE. 0) GO TO 11
   WRITE (11) ((G(I, J), I = 1, NOS), J = 1, NOS)
   IF (IGMAT .NE. 1) GO TO 13
   NN = NOS
   KK = 0
   WRITE (6, 50)
   .50 FORMAT (/1X, *G(I, J) J ACROSS, I DOWN*,/1)
12 LL = KK + 1
   KK = KK + LL
   KK = MINO(KK, N)
   WRITE (6, 300)
300 FORMAT (1X)
   WRITE (6, 20) (I, I = LL, KK)
   WRITE (6, 300)
20 FORMAT (1HO, AX, 11(13, 8X))
   WRITE (6, 300)
   DO 30 I = 1, NN
30 WRITE (6, 40) I, (G(I, J), J = LL, KK)
40 FORMAT (14, 2X, I1(F10.6, 1X))
   IF (KK .LT. NN) GO TO 12
   WRITE (6, 300)
   WRITE (6, 300)
13 CONTINUE
   GO TO 15
11 READ I11 (G(I, J), I = 1, NOS), J = 1, NOS
15 CONTINUE
   RETURN
END
SUBROUTINE AMAT
COMMON C, NOS, NNOS, IRUN, JRUN, NRUN, IB, U, UL, UU, DU, IMAT, IGMAT, XME, IPNCH, IPLT, KF
COMMON/IDET/ID1(140), BSUB1(140)
COMMON/Pol/At1(26), At2(26), Up(6)
COMPLEX Al(140,140), ARI
DIMENSION G(140,140)
EQUIVALENCE (C(1),G(1,1))
EQUIVALENCE (C(1),A(1,1))
DIMENSION S(2)
EQUIVALENCE (S(1), ARI), (S(1), AR), (S(2), AI)
DO 1 K=1,NOS
I=NOS+1-K
DO 1 L=1,NOS
J=NOS+1-L
AIJ=GIJ,JI
1 CONTINUE
DO 2 I=1, NOS
K=IDAI(I)
AR=AT1(K)
AI=AI1(K)
ARI=1./ARI
2 CONTINUE
IF( IMA, EQ, 1. ) GO TO 10
CALL INVERT (A,NOS)
RETURN
10 CONTINUE
WRITE (6,11)
11 FORMAT (1X)
WRITE (6,12)
12 FORMAT (1X,*A(1,J) BEFORE INVERSION  J ACROSS, I DOWN*)
CALL CMOUT
CALL INVERT (A,NOS)
WRITE (6,11)
WRITE (6,13)
13 FORMAT (1X,*A(I,J) AFTER INVERSION  J ACROSS, I DOWN*)
CALL CMOUT
RETURN
END
SUBROUTINE UWSHPE
COMMON C(39200)
COMMON /MISC/NOS, NNOS, IRUN, JRUN, NRUN, IB, U, UL, UU, DU, IMAT, IGMAT, XMER,
1IPNCH, IPLT, KF
COMMON /EEOS/ XE(140), YE(140), ZE(140)
DIMENSION E(3, 3), T(3, 3), TR(3, 3), W(3), EI(3), EJ(3)
COMPLEX A(140, 140)
EQUIVALENCE (C(1), A(1, 1))
DIMENSION S(2)
COMPLEX CI
EQUIVALENCE (S(1), CI), (S(1), AR), (S(2), AI)
DO 3 K=1, 3
DO 3 L=1, 3
TI(K, L)=0.
3
3 TR(K, L)=0.
DO 1 I=1, NOS.
EI(1)=XE(I)
EI(2)=YE(I)
EI(3)=ZE(I)
DO 1 J=1, NOS.
CI=AI(I, J)
EJ(1)=XE(J)
EJ(2)=YE(J)
EJ(3)=ZE(J)
DO 2 K=1, 3
DO 2 L=1, 3
T=EI(K)*EJ(L)
TI(K, L)=TI(K, L)+T*AI
TR(K, L)=TR(K, L)+T*AR
2 CONTINUE
1 CONTINUE
WRITE (6, 200)
200 FORMAT (1H+, 9X, *EX*)
IF (IMAT.EQ.1) CALL TPRINT(TI)
CALL DIAG (TI, E, W, 3)
V=-6.8727*U
CALL PRINT (W, E, V, 1)
RETURN
END
SUBROUTINE RTSHPE
COMMON (C19200)
COMMON/MISC/NNDS,NNDS,JRUN,JRUN,NRUN,IB,U,UL,UU,DU,IMAT,IGMAT,XMR,
IIPNCH,IPLT,KF
COMMON/ROSC/ XR(140),YR(140),ZR(140)
COMMON /EEOS/ XE(140),YE(140),ZE(140)
COMMON /EMSOS/XM(140),YM(140),ZM(140)
COMMON/IDET/IDA(140),BSUBI(140)
DIMENSION E(3,3),T1(3,3),TR(3,3),W(3),EI(31),EJ(31),EC(31),EM(31),
EI(31),TI(3,3),TRM(3,3),TP(3,3),TRP(3,3)
COMPLEX A(140,140)
EQUIVALENCE (C11,A11)
DIMENSION S(2)
COMPLEX CI
EQUIVALENCE (S(1),CI1),(S(1),AR),(S(2),AI)
DO 5 K=1,3
DO 5 L=1,3

t(K,L)=0.

t(K,L)=0.

t(K,L)=0.

t(K,L)=0.

5 TP(K,L)=0.

NK=0
DO 1 K=1,10
DO 2 L=1,K
D(1)=XR(L)-XR(K)
D(2)=YR(L)-YR(K)
D(3)=ZR(L)-ZR(K)
E(11)=XE(L)
E(12)=YE(L)
E(31)=ZE(L)
C1=A(K,L)
IF(K.NE.L) GO TO 10
AI=AI/2.
AR=AR/2.

10 CONTINUE
EJ(11)=XE(L)
EJ(12)=YE(L)
EJ(31)=ZE(L)
EM(11)=XM(L)
EM(12)=YM(L)
EM(31)=ZM(L)
B=BSUBI(L)*4
DO 6 II=1,3
M=II/3
KK=II+1-M*3
M=II+1/3
LL=II+2-M*3
EC(II)=EI(KK)*EJ(II)-EI(II)*EJ(KK)
DO 7 JJ=1,3
Tp=EC(II)*D(JJ)
TM=EI(II)*EM(JJ)*B
TIP(I1,JJ)=TIP(I1,JJ)+TP*AI
TIM(I1,JJ)=TIM(I1,JJ)+TM*AI
TRP[II, JJ] = TRP[II, JJ] + TP*AR
7 CONTINUE
6 CONTINUE
5 CONTINUE
4 CONTINUE
3 CONTINUE
2 CONTINUE
1 CONTINUE
DO 11 II = 1, 3
DO 11 JJ = 1, II
TPI[IJ, II] = TPI[IJ, II]
TIM(II, JJ) = TIM(II, JJ)
TRP(IJ, II) = TRP(IJ, II)
TRM(IJ, II) = TRM(IJ, II)
TI(I, JJ) = TPI[IJ, II] + TIM(IJ, II)
TR(I, JJ) = TRP(IJ, II) + TRM(IJ, II)
TR(IJ, II) = TRP(IJ, II)
11 CONTINUE
IF (IB.EQ.0) GO TO 8
WRITE (6, 200)
IF (IMAT.EQ.1) CALL TPRINT(TIP)
V = U*U*0.00043167
CALL DIAG (TIP, E, W, 3)
CALL PRINT (W, E, V, 0)
WRITE (6, 201)
IF (IMAT.EQ.1) CALL TPRINT(TIM)
CALL DIAG (TIM, E, W, 3)
CALL PRINT (W, E, V, 0)
WRITE (6, 202)
IF (IMAT.EQ.1) CALL TPRINT(TRP)
V = -V*3000.0
CALL DIAG (TRP, E, W, 3)
CALL PRINT (W, E, V, 0)
WRITE (6, 203)
IF (IMAT.EQ.1) CALL TPRINT(TRM)
CALL DIAG(TRM, E, W, 3)
CALL PRINT (W, E, V, 0)
200 FORMAT (1H+, 9X, *CP*)
201 FORMAT (1H+, 9X, *CM*)
202 FORMAT (1H+, 9X, *RP*)
203 FORMAT (1H+, 9X, *RM*)
8 WRITE (6, 206)
206 FORMAT (1H+, 9X, *CD*)
IF (IMAT.EQ.1) CALL TPRINT(TI)
V = U*U*0.00043167
CALL DIAG (TI, E, W, 3)
CALL PRINT (W, E, V, 0)
RETURN
END
SUBROUTINE PRINT(W,E,V,NID)
COMMON/MISC/NOS,NNOS,IRUN,JRUN,NRUN,JRUN,IB,UL,UU,DU,IMAT,IGMAT,XMER,
1IPNCH,IPLT,KF
COMMON/WWW/WWW(101,4)
DIMENSION W(31),E(3,31)
DIMENSION Z(31)
WW=0.
DO 1 I=1,3
W(I)=W(I)*V
WW=WW+W(I)
1 DO 1 J=1,3
IF (ABS(E(I,J)) .GT. 1.) E(I,J)=1.
 1 E(I,J)=ACOS(E(I,J))*57.3
WW=WW/XMER
IF(NID.EQ.0) GO TO 2
WWW(KF,NID)=WW
2 CONTINUE
WRITE (6,200) WWW,((W(J),E(I,J),I=1,3),J=1,3)
200 FORMAT (12X,F12.3,3X,31F12.3,2X,3F6.1)
DO 3 I=1,3
Z(I)=0.
DO 4 J=1,3
 4 Z(I)=Z(I) + (3.*W(J)*((COS(.0174533*E(I,J)))**2)/XMER)
3 CONTINUE
ZZ=.5*(Z(1)+Z(2))
WRITE (6,210) Z(1),ZZ,Z(1),Z(2)
210 FORMAT (130X,'PARA=',F10.3,' PERP=',F10.3,' XX=',F10.3,' YY=',
1,F10.3)
300 FORMAT (11.3,3(F8.2,3F5.1))
RETURN
END

SUBROUTINE CMAOUT
COMMON C(39200)
COMMON/MISC/NOS,NNOS,IRUN,JRUN,NRUN,JRUN,IB,UL,UU,DU,IMAT,IGMAT,XMER,
1IPNCH,IPLT,KF
COMPLEX A(140,140)
EQUIVALENCE A(1140,140)
N=NOS
K=0
1 L=K+1
K=K+7
K=MINO(K,N)
WRITE(6,200)
200 FORMAT (1X)
WRITE (6,2) (I,1=1,L,K)
2 FORMAT (1H0,11X,7(I3,14X))
WRITE(6,200)
DO 3 I=1,N
3 WRITE (6,4) I,(A(I,J),J=1,L,K)
4 FORMAT (14,2X,7(2F8.4,1X))
IF(K.LT.N) GO TO 1
WRITE (6,200)
WRITE (6,200)
RETURN
END
SUBROUTINE TPRINT (T)
DIMENSION T(3,3)
WRITE (6,201)
201 FORMAT (1X)
DO 1 I=1,3
1 WRITE (6,200) (T(I,J),J=1,3)
200 FORMAT (1X,3F10.4)
WRITE (6,201)
RETURN
END

SUBROUTINE INVERT (T,N)
COMPLEX T(140,140),R(140),S(140),D
DO 100 L=1,N
D=1./T(L,L)
DO 30 I=1,N
IF (L-I) 15,20,25
15 R(I)=T(I,L)*D
S(I)=T(I,L)
GO TO 30
20 R(I)=-D
GO TO 30
25 R(I)=T(L,I)*D
S(I)=T(L,I)
30 CONTINUE
DO 100 I=1,N
DO 100 J=1,N
IF (I-L) 35,50,35
35 IF (J-L) 40,45,40
40 T(I,J)=T(I,J)-R(I)*S(J)
GO TO 100
45 T(I,J)=R(I)
GO TO 100
50 T(I,J)=R(I)
100 CONTINUE
DO 200 I=1,N
DO 200 J=1,N
T(I,J)=-T(I,J)
200 T(J,I)=T(I,J)
RETURN
END
SUBROUTINE DIAG (A,EIVR,E,N)
DIMENSION A(3,3),EIVR(3,3),E(3)
IEGEN=0
IF(N-1 .GT. 2,2,1)
 2 EIVR(1,1)=1.0
E(1)=A(1,1)
RETURN
1 IF(IEGEN) 102,99,102
99 DO 101 J=1,N
DO 100 I=1,N
100 EIVR(I,J)=0.0
101 EIVR(I,J)=1.0
DO 888 I=1,N
DO 888 J=1,N
IF (ABS(A(I,J)).GT.10E-12) GO TO 887
888 CONTINUE
GO TO 134
887 CONTINUE
C FIND THE ABSOLUTELY LARGEST ELEMENT OF A
102 ATOP=0.
DO 111 I=1,N
DO 111 J=1,N
IF (ATOP-ABS(A(I,J))111,111,111
104 ATOP=ABS(A(I,J))
111 CONTINUE
IF(ATOP)109,109,113
109 RETURN
C CALCULATE THE STOPPING CRITERION -- DSTOP
113 AVGF=FLOAT(N*(N-1))*10.0
DO 114 JJ=2,N
DO 114 II=2,II
S=A(II-1,JJ)/ATOP
114 D=S*S+D
OSTOP=11.E-06*D
C CALCULATE THE THRESHOLD, THRSH
C THRSH = SQRT(D/AVGF)*ATOP
C START A SWEEP
C
115 IFLAG=0
DO 130 JCOL=2,N
JCOL1=JCOL-1
DO 130 IROW=1,JCOL1
AIJ=A(IROW,JCOL)
C COMPARGE THE OFF-DIAGONAL ELEMENT WITH THRSH
C IF(ABS(AIJ)-THRSH)130,130,117
117 AIJ=A(IROW,IROW)
AJJ=AJ(JCOL,JCOL)
S=AJJ=AI

CHECK TO SEE IF THE CHOSEN ROTATION IS LESS THAN THE Rounding ERROR
IF SO, THEN DO NOT ROTATE.

IF (ABS(AIJ)-1.0EQABS(S)) .LT. 30.0, 118

IFLAG=1

IF THE ROTATION IS VERY CLOSE TO 45 DEGREES, SET SIN AND COS
TO 1/IRQOT 21.

IF (1.0 .EQ 10*ABS(AIJ)-ABS(S)) .LT. 116, 119, 119

S=.70710678118655
C=S
GO TO 120

CALCULATION OF SIN AND COS FOR ROTATION THAT IS NOT VERY CLOSE
TO 45 DEGREES

T=AIJ/S
S=0.25/SORT(0.25+T*T)
COS=C
SIN=S
C=SORT(0.5+S)
S=2.*T*S/C

CALCULATION OF THE NEW ELEMENTS OF MATRIX A

DO 121 I=1,IROW
T=AI(I,IROW)
U=AI(I,JCOL)
AI(I,IROW)=C*T-S*U
121 AI(I,JCOL)=S*T+C*U
I2=IROW+2
IF (J-JCOL) .EQ. 127, 127, 123
CONTINUE

DO 122 J=JCOL,IROW
T=AI(IROW,J)
U=AI(JCOL,J)
AI(IROW,J)=C*T-S*U
122 AI(JROW,J)=C*U-S*T
123 AI(JCOL,J)=S*AIJ+C*AJJ
AI(IROW,IROW)=C*AI(ROW,IROW)-S*(C*AIJ-S*AJJ)
DO 124 J=JCOL,N
T=AI(IROW,J)
U=AI(JCOL,J)
AI(IROW,J)=C*T-S*U
124 AI(JCOL,J)=S*T+C*U

ROTATION COMPLETED.

SEE IF EIGENVECTORS ARE WANTED BY USER

IF (IEGEN) 126, 131, 126
DO 125 I=1,N
T = EI VRI(I, JCOL) + C* T - EI VRI(I, JCOL)* S
125  E I V P(I, JCOL) = S* T - EI VRI(I, JCOL)* C

C   CALCULATE THE NEW NORM D AND COMPARE WITH DSTOP
C
126  CONTINUE
   S = AIJ/ATOP
   D = D - S* S
   IF(D < DSTOP) GO TO 1260
129  TIRESH = SQRT(0/DATAGF*ATOP
130  CONTINUE
   IF(IFLAG .Lt. 115) GO TO 134
134  DO 998 I = 1, N
998  E(I) = A(I, I)
   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
C  DIMENSION X(NPTS), Y(NPTS), IGRID(105), XAXIS(11)
C
C  INTEGER BLANK, DOT, STAR, IGRID, PLUS
C  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 (+, I, 5H PTS ) )
905  FORMAT(16X, 11(1H+, 9X+) )
9800 FORMAT(46H SCALING 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.103.0.OR.JZERO.LT.) 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
DO 11 J=1,105
11 IGRID(J)=BLANK
15 IGRID(IZERO)=PLUS
IGRID(104)=DOT
IGRID(2)=DOT
DO 12 K=1,NPTS
TEST=(YMAX-Y(K))/YINCR*1.5
IF(TEST.NE.1) 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-11-1*YINCR
IF(ABS(YAXIS).LT.YAXMIN) YAXIS=0.
PRINT 902,YAXIS,(IGRID(J),J=1,105)
CONTINUE
PRINT 903
PRINT 905
DO 20 M=1,11
XAXIS(M)=XMAX-XINCR*(FLOAT(11-M))*10.0
IF(ABS(XAXIS(M)).LT.XAXMIN) XAXIS(M)=0.
CONTINUE
PRINT 904,XAXIS,NPTS
RETURN
PRINT 9800
CALL EXIT
END

SUBROUTINE PLSCALCV,VMAX,VINCR,NPTS,NOIVIS)
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*VINC
C DIMENSION VINPTS
C
VMIN=V(I)
VMAX=V(I)
DO 10 I=1,NPTS
   IF(V(I).LT.VMIN) VMIN=V(I)
   IF(V(I).GT.VMAX) VMAX=V(I)
   QANGE=VMAX-VMIN
10 CONTINUE
   IF(QANGE.EQ.0) GO TO 8000
   QANGE=0.4342944*ALOG(QANGE)
   IF(QANGE.EQ.0) GO TO 30
   IRANGE=QANGE
   GO TO 40
30
   IRANGE=-QANGE
   IRANGE=-IRANGE-1
40
   QANGE=QANGE-FLOAT(IRANGE)
   RANGE=10.* QANGE
C RANGE IS BETWEEN 1.0 AND 10.0
C
43  IF(QANGE.GT.2.5) GO TO 41
    RANGE=2.5
    GO TO 50
41  IF(QANGE.GT.5.0) GO TO 42
    RANGE=5.0
    GO TO 50
42  RANGE=10.0
50  QANGE=RANGE*(10.* QANGE)
C TRANGE IS NOW 2.5, 5.0, OR 10.0 TIMES A POWER OF TEN
C
51  IMAX=VMAX/FLOAT(5.0*QANGE)
52  XMAX=5.0*QANGE*FLOAT(IMAX+1)
53  IF(VMIN.GT.XMAX-QANGE) GO TO 100
54  RANGE=RANGE/10.* QANGE+1
55  IRANGE=IRANGE+1
56  GO TO 43
57  IMAX=-VMAX/(5.0*QANGE)
58  XMAX=5.0*QANGE*FLOAT(-IMAX+1)
59  IF(VMIN.GT.XMAX+QANGE) GO TO 100
60  RANGE=RANGE*2.0
61  IF(RANGE.GT.10.) GO TO 43
62  IRANGE=IRANGE+1
63  GO TO 43
100  VMIN=XMAX
VMAX=XMAX+QANGE
RETURN
8000 PRINT 9800
9800 FORMAT(45H1PLSCAL CALLED TO SCALE ARRAY WITH ZERO RANGE)
CALL EXIT
END
II. **PROGRAM ROTOPR.** Revised version of ROTOPM designed to change $\theta$ a number of times for up to 4 oscillators after $A_{ij}$ has been inverted and stored at each freq. (That is, $G_{ij}$ determined from a given set of $q$ is not changed.)

Altered subroutines: INPUT, FREQU, PRINT, UVSHPE, RTSHPE

New subroutines: NEWMU, NEWDIR

New/Common/: ROTATE, VECT, MNOS, (MISC changed)

New restrictions:
- Calculates 101 frequency points but currently handles only 30 freq points in punch and plot.
- Plots and punches CD only.
- Can change $\theta$ for 4 different oscillators.
- Each of these 4 osc's can be at 16 positions on helix.
- Each of these 4 osc's can be given a max of 10 new $\theta$ values but the total number of spectra calculated must not be greater than 40.
- Only seven osc's can be specified on IBASE card.

Reference vector in input deck defined as:
Cartesian coords of unit vector for $C_6+\overline{C_5}$ pyrimidines
$C_4+\overline{C_5}$ purines

where atom 2 + atom 1 is equal to

$$\left[ \frac{x_1 - x_2}{|R|}, \frac{y_1 - y_2}{|R|}, \frac{z_1 - z_2}{|R|} \right]$$

$$|R| = \left( (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 \right)^{1/2}$$

**ROTOPR INPUT DECK**— same as ROTOPM with following changes:

Reference base deck: after each TP, PP (perpendicular unit vector) card, add 1 card specifying unit vector for Reference vector $x, y, z$ format(3F10.3)
Polymer deck:

New IBASE card:
IBASE,NBR, (IOSC 1→7), NNBR
FORMAT(A1,3X,I2,4X,7(I2,3X),27X,I1)
IBASE=ID of ref. base as before
NBR=# of unchanged osc's
IOSC specifies 1→NBR ref. osc's
NNBR=# of osc for which θ is changed

Position card (POS) defined as before, but now 16 positions (Format 16F5.0) may be specified at once

New cards, 1 for each NNBR osc., immediately following POS card:
IOSC,NX,TDEVOE (1→10)
FORMAT (I1,1X,I2,2X,10E7.0)

IOSC=identifies ref osc by #
NX=# of θ given to this osc
TDEVOE=θ(1→NX) in degrees

Note: NNBR may be 0 for some or all IBASE cards
Main program of ROTOPR - same as ROTOPM

SUBROUTINES slightly altered from ROTOPM, to set up data for NEWNU and for new print format:

- INPUT
- FREQU
- UVSHPE
- RTSHPE
- PRINT

SUBROUTINES same as ROTOPM:
- ASHAPE
- INTRAC
- AMAT
- CMAOUT
- TPRINT
- INVERT
- DIAG
- FENTHR
- FLSCAL

SUBROUTINE NEWNU
COMMON/MISC/ROS,MNOS,JRUN,IRUN,MRUN,IB,UL,DU,IMAT,IGMAT,XMER,
1IPNCH,IPJL,KF,MRUN.
COMMON/MNOS/MNOS,IDT(4),NPS(4),NX(4),TDEVOE(4,10)
MRUN=0
IF(MNOS.EQ.0) GO TO 54
MN=MNOS
IEND=MX(MN)
DO 50 MI=1,IEND
DBASE=DT(MN,MI)
NBRPOS=NPS(MN)
IF(IRUN.NE.0) GO TO 70
WRITE(6,100) MN,TDEVOE(MN,MI)
50 CONTINUE
CALL NEWDIR(MN,IOBASE,NBRPOS,DIR)
IF(MNOS-1) GT.01 GO TO 60
MRUN=MRUN+1
CALL UVSHPE
CALL RTSHPE
GO TO 50
60 CONTINUE
MN1=MNOS-1
JEND=MX(MN1)
DO 51 MJ=1,JEND
DBASE=DT(MN1)
NBRPOS=NPS(MN1)
IF(IRUN.NE.0) GO TO 71
WRITE(6,100) MN1,TDEVOE(MN1,MJ)
51 CONTINUE
71 CONTINUE
    CALL NEWDIR(MN1, IOBASE, NBRPOS, DIR)
    IF(MNOS-2 .GT. 0) GO TO 61
    MRUN = MRUN + 1
    CALL UVSHPE
    CALL RTSHPE
    GO TO 51
61 CONTINUE
    MN2 = MNOS - 2
    KEND = NX(MN2)
    DO 52 MK = 1, KEND
        DIR = -TDEVÖE(MN2, MK)
        IOBASE = IDT(MN2)
        NBRPOS = NPS(MN2)
        IF(IRUN .NE. 0) GO TO 72
        WRITE (6, 1001) MN2, TDEVÖE(MN2, MK)
    52 CONTINUE
    CONTINUE
    CALL NEWDIR(MN2, IOBASE, NBRPOS, DIR)
    IF(MNOS-3 .GT. 0) GO TO 62
    MRUN = MRUN + 1
    CALL UVSHPE
    CALL RTSHPE
    GO TO 52
62 CONTINUE
    MN3 = MNOS - 3
    LEND = NX(MN3)
    DO 53 ML = 1, LEND
        MRUN = MRUN + 1
        DIR = -TDEVÖE(MN3, ML)
        IOBASE = IDT(MN3)
        NBRPOS = NPS(MN3)
        IF(IRUN .NE. 0) GO TO 73
        WRITE (6, 1001) MN3, TDEVÖE(MN3, ML)
    53 CONTINUE
    CONTINUE
    CALL NEWDIR(MN3, IOBASE, NBRPOS, DIR)
    CALL UVSHPE
    CALL RTSHPE
53 CONTINUE
52 CONTINUE
51 CONTINUE
50 CONTINUE
54 MRUN = 1
    CALL UVSHPE
    CALL RTSHPE
55 RETURN
END
SUBROUTINE NEWDIR(IN, ID, NP, IDIR)
COMMON/MISC/NO, NGOS, IRUN, JRUN, NRUN, IB, U, UL, UU, DU, IMAT, IGMAT, XMER,
1 IPNC, IPLT, KF, MRUN
COMMON/ROTA/THE, STR(4, 20), PS(4, 20), IONOS(4, 20)
COMMON/VECT/ P(4, 3), E(4, 3)
COMMON/VEOS/XE(1401), YE(1401), ZE(1401)
ST=SIGN(0.1745329)*DIRI
CT=COS(ST)/SIGN(0.1745329)*DIRI
TX=ST*(E(1D, 2)*P(1D, 3)-E(1D, 3)*P(1D, 2))+E(1D, 1)*CT
TY=ST*(E(1D, 3)*P(1D, 1)-E(1D, 1)*P(1D, 3))*E(1D, 2)*CT
TZ=ST*(E(1D, 1)*P(1D, 2)-E(1D, 2)*P(1D, 1))*E(1D, 3)*CT
T=ACOS(TZ)
IF(ABS(T) .LE. 10**(-6)) GO TO 3
PHI=ACOS(TX/SINI)
IF(TY) 2, 3, 4
2 PHI=-PHI
GO TO 4
3 PHI=0.
4 CONTINUE
T=T*57.2958
PHI=PHI*57.2958
DO 1 I=1, NP
TT=T
IF(STR(N, 1) .EQ. (-1.)) TT=180.-T
789 FORMAT (1DX, *X*)
PP=(PHI*STR(N, 1))+PS(N, 1)*THE
K=IONOS(N, 1)
XE(K)=SINI*TT*0.17453291*COS(PP*0.1745329)
YE(K)=SINI*TT*0.17453291*SIN(PP*0.1745329)
ZE(K)=COS(1TT*0.1745329)
IF(IRUN .NE. 0) GO TO 5
WRITE(6, 100) XE(K), YE(K), ZE(K)
100 FORMAT (2DX, 3(F6.3, 3X))
5 CONTINUE
1 CONTINUE
RETURN
END
III. **PROGRAM QMCD**

Calculates CD and absorption by the matrix method. Units for all terms of the Hamiltonian are kK. Subroutines INPUT and INTRAC are modified versions of the corresponding subroutines in ROTOPM.

Input deck: same as for ROTOPM with following changes:

(a) no polarizability deck is used

(b) $\Omega_1$ (in kK) for each reference base oscillator must be punched in columns 73-77 on card specifying position and direction of each transition dipole. ROTOPM does not read this space, so the same cards can be used for both programs.

(c) in polymer deck, a new card must be inserted after title card:

```
FU,FD,DELTA,NHW,(HWIDTH(I),I=1,10)
Format(2F10.2,F5.2,I2,3X,10FS.2)
```

FU = max freq. to be calculated
FD = min freq. to be calculated
DELTA = freq. increment
NHW = number of different halfwidths for which calculation is to be done, \( \leq 10 \).
HWIDTH = halfwidth, in kK (1 to NHW)

(d) IGMAT option of ROTOPM is now IHMAT option for printing of H matrix
PROGRAM QMCD (INPUT, OUTPUT, PUNCH, TAPE5=INPUT, TAPE6=OUTPUT,
1 TAPE7=PUNCH)
COMMON H(9870), COMMON/MISC/ NOS, NNOS, NRUN, FU; FD, DELTA, NHW, NWIDTH(10), IHMAT, XMER,
1 NPLCH, IPLT
COMMON/ROSC/ XR(140), YR(140), ZR(140), F1(140)
COMMON/EOOS/ XE(140), YE(140), ZE(140)
COMMON/EMOS/ XM(140), YM(140), ZM(140)
COMMON/OP/ P1(140), D1(140), F1(140)
DIMENSION CC(19600)
NRUN=0
1 CALL INPUT
DO 2 I=1, NOS
J=(I*(I-1))/2+1
H(IJ)=F1(I)
2 CONTINUE
IF(IHMAT,EQ.0) GO TO 10
WRITE(6,20)
20 FORMAT (1X)
WRITE (6,21)
21 FORMAT (/1X,*HAMILTONIAN MATRIX (IN KK) / DIAGONAL ELEMENTS ARE FR
1 IEO, OFF DIAGONALS ARE V(J,JI)*)
CALL TRIMAT(NH,NOS)
10 CONTINUE
CALL DIAG21H(CC, F, NOS)
WRITE (6,22)
22 FORMAT (/1X,*EIGENVECTORS / POLYMER TRANS ACROSS, MONOMER TRANS DO
1 INN*)
CALL QUAMACCC, NOS)
WRITE (6,30)
30 FORMAT (/1X,*UNLABELED NUMBERS BELOW ARE X,Y,Z COMPONENTS OF MU
1 SUBK THEN N SUBK*,//,IX,*RESPECTIVELY COMING FROM MONOMER OSC 1 TO
1 INOS*)
WRITE (6,31)
31 FORMAT (/1X,*POLYMER TRANS*,3X,*FREQ*,20X,*PR*,20X,*DO*,13X,*XMU*,
1 P*,*YM*,7X,*ZMU*,7X,*XMG*,6X,*YMAG*,6X,*ZMAG*)
DO 3 I=1, NOS
SX=0.
SY=0.
SZ=0.
TX=0.
TY=0.
TZ=0.
M=(I(I-1)/2)+1
WRITE (6,20)
DO 4 J=1, NOS
N=M+J-1
XJ=CC(N)*XE(J)
SX=SX+XJ
YJ=CC(N)*YE(J)
SY=SY+YJ
ZJ=CC(N)*ZE(J)
SZ=SZ+ZJ
X2=CC(N)*YR(J)+ZE(J)-YE(J)*ZR(J)+F1(I)
Y2=CC(N)*ZR(J)+XE(J)-ZE(J)*XR(J)+F1(I)
Z2=CC(N)*XR(J)+YE(J)-XE(J)*YR(J)+F1(I)
TX = TX + X2
TY = TY + Y2
TZ = TZ + Z2
WRITE (6,31) X1, Y1, Z1, X2, Y2, Z2
31 FORMAT (20X, 6(F16.6, 1X))
4 CONTINUE
TX = TX * 3.14159 * (10. ** (-5.1))
TY = TY * 3.14159 * (10. ** (-5.1))
TZ = TZ * 3.14159 * (10. ** (-5.1))
A1 = SX * TX
A2 = SY * TY
A3 = SZ * TZ
WRITE (6, 32) A1, A2, A3
32 FORMAT (/40X, 3F15.6)
R(I) = (SX * TX) + (SY * TY) + (SZ * TZ)
D(I) = (SX * SX) + (SY * SY) + (SZ * SZ)
WL = 10000. / F(I)
WRITE (6, 20)
WRITE (6, 24) I, F(I), WL, R(I), D(I), SX, SY, SZ, TX, TY, TZ
3 CONTINUE
CALL SHAPE(10)
NRUN = 1
GO TO 1
END

SUBROUTINE TRIMAT(A,N)
DIMENSION A(I)
II = 0
6 IK = II + 1
WRITE (6, 1)
WRITE (6, 1)
1 FORMAT (1H1)
II = II + 1
IL = MINO(13, N)
WRITE (6, 3) (I, I = IK, IL)
3 FORMAT (9X, 14(I3, 5X))
WRITE (6, 1)
IJ = (II / 2) * IK + 1
IH = IK
DO 4 I = IH, N
IJ = IJ + 1
I = I - 1, N
4 WRITE (6, 5) I, (A(IL), IL = IJ, IK)
5 FORMAT (1X, I3, 2X, 14F8.4)
II = II + 1
END
7 RETURN

SUBROUTINE INPUT: as in ROTOPM but
1) reads FMAX(J,I) in FORMAT 304
2) adds F(NOS) = FMAX(K,I) for each oscillator

SUBROUTINE INTRAC
COMMON HI(9870)
COMMON/MISC/ NOS,NNOS,NRUN,FU,FD,DELTA,NHM,HWIDTH(10),IHMAT,XMER,
IPNCH,IPLT
COMMON/MPOL/ Q(22,7,4),D(7,4),D1(140),IDDB(140),IDQ(140)
COMMON/ATCMS/ XA(22,140),YA(22,140),ZA(22,140)
DIMENSION R(3)
IF(NNOS.GT.0) GO TO 26
DO 25 I=1,NOS
JI=(I*(I-1))/2
DO 25 J=1,I
II=J+J
25 H(I,J)=0.
RETURN
26 CONTINUE
NN=NNOS+1
DO 1 I=NN,NOS
II=IDB(I)
JJ=IDQ(I)
JI=(I*(I-1))/2
DO 2 J=1,NNOS
II=J+J
JJ=IDR(J)
LL=IDQ(J)
GG=0.
DO 3 K=1,LL
DO 4 L=1,LL
R(1)=XAX(K,I),XAL(J)
R(2)=YAX(K,I),YAL(J)
R(3)=ZAX(K,I),ZAL(J)
DD=0.
DO 5 M=1,3
DO=DD+R(M)*R(M)
D=DSRT(DD)
M=IDBB(JJ)
N=IDDB(JJ)
GG=GG+((O(K,M,II)*Q(L,N,JJ)))/D
4 CONTINUE
3 CONTINUE
HI(I,J)=(4*80298*4,80298)/(2.9979*6.625611*GG*100.
HI(I,J)=HI(I,J)/2.
2 CONTINUE
DO 1 J=NN,1
II=J+J
HI(I,J)=0.
1 CONTINUE
RETURN
END
SUBROUTINE SHAPE(N)
COMMON/MISC/NOS,NNOS,NNRUN,FU,FD,DELTA,NHW,NWIDTH(10),IXMAT,IXMER,
                   1PINC,PLT
COMMON/OPT/R(140),D(140),F(140)
DIMENSION CC(101),EXT(101),FREQ(101)
DIMENSION GAUS(140,76)
FUNC1(IA,UA,SA,UA1=DA*UA/SA*EXP(-.69*((UA-IA)/SA)**2))
FUNC2(RA,UA,SA,UA1=RA*UA/SA*EXP(-.69*((UA-IA)/SA)**2))
DO 5 I=1,NHW
      WRITE (6,100) NWIDTH(I)
100 FORMAT (///,1X,*OPTICAL DATA CALCULATED FROM GAUSSIANS WITH HALF WIDTH=*
      F10.2,* KK=*
      WRITE (6,101) XMER
101 FORMAT (1X,*POLYMER DATA HAS BEEN DIVIDED BY NUMBER OF MONOMERS =*
      F4.0)
      WRITE (6,102)
102 FORMAT (///,1X,*FREQ*,10X,*WVLTH*,21X,*EXT COEFF*,21X,*CDI(DELTA EXT)*)
      DO 3 I=1,F4,0
      CONST=51.15
      NPTS=((FU-FD)/DELTA1+1.01
      SA=NWIDTH(I)
      DO 1 I=1,NPTS
         A=0.
         B=0.
         UA=FD+((I-1)*DELTA)
         FREQ(I)=UA
         DA=D(I)
         RA=R(I)
         UA=F(I)
         A=A+FUNC1(DA,UA,SA,UA)
         GAUS(I)=FUNC2(RA,UA,SA,UA)
         B=B+GAUS(I,1)
      2 CONTINUE
      EXT(I)=A*CONST/XMER
      CDI(I)=B*4.*CONST/XMER
      WVLTH(I)=10.**4.1/UA
      WRITE (6,103) UA,WVLTH,EXT(I),CDI(I)
103 FORMAT (1X,F5.7,9X,F6.2,20X,F8.2,23X,F7.2)
      1 CONTINUE
      WRITE (6,110)
110 FORMAT (///,1X,*TERM PROPORTIONAL TO CD CONTRIBUTION OF EACH POLYMER O*
      1SC 1 TO 140*)
      NI=0
17 CONTINUE
      NJ=NI+1
      NK=NI+10
      N=MINO(NK,NPTS)
      DO 18 J=1,N
         WRITE (6,117) J,GAUS(J,JN),JN=NJ,NK
117 FORMAT (1X,13,10F13.4)
      18 CONTINUE
      NI=NI+10
      IF(NI,LT,NPTS) GO TO 17
      IF(IPLT.EQ.0) GO TO 3
      WRITE (6,9001
SUBROUTINE OUAMA (A, NCOL)
DIMENSION A(1)
NROW=NCOL
ID=NROW-NCOL
II=0
2. IK=II+1
WRITE (6,1)
WRITE (6,2)
1 FORMAT (1H )
II=II+1
IL=MINO(II,NCOL)
IKK=IK+ID
ILL=IL+ID
WRITE (6,3) (I, I=II, I=IKK, ILL)
3 FORMAT (10X,14(13,5X))
WRITE (6,1)
INROW=NROW-MINO(13,(NCOL-IK))
IJ=II*INROW
IK=IJ+INROW
DO 4 I=1, NROW
IJ=IJ+1
IK=IK+1
4 WRITE (6,5) I, A(I), IL=IJ, IK, NROW)
II=II+1
IF (II-NCOL) 2, 6, 6
5 FORMAT (1X,13,3X,14F8.4)
6 RETURN
END
SUBROUTINE DIAG2 (A, R, E, N)
DIMENSION A(1), R(1), E(1)
IQ=N*N
DO 15 I=1, IQ
15 R(I)=O.
IQ=N
DO 20 I=1, N
IQ=IQ+N+1
R(I)=1.
20 CONTINUE

ANORM=0.
DO 35 I=1, N
DO 35 IM=1, N
IF (I-IM) 30, 35, 30
30 IL=I+(I*M+M-M)/2
ANORM=ANORM+A(IL)*A(IL)
35 CONTINUE

ANORM=SQRT(ANORM)
AKK=N
ANRMX=ANORM*1.0E-12/AKK
IND=0
THR=ANORM
45 THR=THR/AKK
50 L=1
55 M=L+1
60 MQ=(M*M-M)/2
LQ=L*L-L/2
LM=L+MQ
IF (ABS(A(LM))>THR) 130, 65, 65
65 IND=1
LL=L+LQ
MM=M+MQ
X=(A(LL)-A(MM))/2.
Y=(A(LM)*SQR(A(LM)+A(LM)+X*X))
IF (X) 70, 75, 75
70 Y=-Y
75 SINX=Y/SQR(1.+(1-SQR(1.-Y*Y)))*SINX
COSX1=1.-SINX2
COSX=SQR(COSX2)
SINCS=SINX*COSX
ILQ=N*(L-1)
IMG=N*(M-1)
DO 125 I=1, N
IO=(I*1-1)/2
IF (I-L) 80, 115, 80
80 IF (I-M) 85, 115, 90
85 IM=I+MQ
GO TO 95
90 IM=I+IQ
95 IF (I-L) 100, 105, 105
100 IL=I+LQ
GO TO 110
105 IL=L+IQ
110 X=A(IL)*COSX-A(IM)*SINX
A(IM)=A(IL)*SINX+A(IM)*COSX
A(I) = X
115 ILR = ILO + 1
IMR = IMQ + 1
X = R(ILR) * COSX - R(IMR) * SINX
R(IMR) = R(ILR) * SINX + R(IMR) * COSX
R(ILR) = X
125 CONTINUE
CONTINUE
X = 2. * A(LM) * SINC
Y = A(LL) * COSX2 + A(MM) * SINK2 - X
X = A(LL) * SINK2 + A(MM) * COSX2 + X
A(LM) = (A(LL) - A(MM)) * SINC * A(LM) * (COSX2 - SINK2)
A(LL) = Y
A(MM) = X
130 IF (M - N) 135, 140, 135
135 M = M + 1
GO TO 50
140 IF (L - (N - 1)) 145, 150, 145
145 L = L + 1
GO TO 55
150 IF (IND) 160, 160, 155
155 IND = 0
GO TO 50
160 IF (THI - ANRMX) 165, 165, 165
165 IQ = -N
DO 185 I = 1, N
IQ = IQ + N
LL = I + (I * I - I - 1)/2
IM = N * (I - 1)
DO 185 IL = I, N
IM = IM + N
MM = IL + (IL * IL - IL)/2
IF (A(LL) - A(MM)) 170, 185, 185
170 X = A(LL)
A(LL) = A(MM)
A(MM) = X
DO 180 LM = 1, N
ILR = IQ + LM
IMR = IM + LM
X = P(ILR)
R(ILR) = R(IMR)
180 R(IMR) = X
185 CONTINUE
DO 3598 I = 1, N
K = I * (I - I - 1)/2 + 1
3598 E(I) = A(K)
RETURN
END

SUBROUTINES PRNPLT and PLSCAL as in ROTOPM
IV. PROGRAM BASES (Author: I. Tinoco, Jr.)

PROGRAM BASES(INPUT, OUTPUT, PUNCH, TAPE5=INPUT, TAPE6=OUTPUT, TAPE7=PUNCH, INCH)

DIMENSION R(201), THEA(201), Z(201), AME(101), X(201, 3), TMAG(201), XX(13, 3)
1, XT(3, 1), THI(101), P16(101), U(3), THE1(101), PHI(101), PERP(3), QX(101)
2, B01(9), BMAG(9), B01(10, 20), QBX(101), BTHI(101), BPHI(110), P(3), Q(3)
3, QBY(T), QZ(T), QY(T), QZ(T)

DIMENSION QR(101), QT(101), QBR(101), QBT(101)

DIMENSION CPP(22)

600 READ(5, 9) (AME(I), I=1, 10)
9 FORMAT(10A8)
WRITE (6, 610) (AME(I), I=1, 10)

610 FORMAT (1H1, 1X, 10A8)

C DEVEO TINOCO NUMBERING SYSTEM
C ORDER OF PURINE ATOMS / N1, C2, N3, C4, C5, C6, N7, C8, C9, (N60RN2), (O6)
C ORDER OF PYRIMIDINE ATOMS / N1, C2, N3, C4, C5, C6, (O2), (O40RN4)

READ(5, 10) M

10 FORMAT (I2)
READ (5, 133) (R(I), I=1, M)

133 FORMAT (3F10.3)

C FIND BEST LEAST SQUARES PLANE THRU POINTS
DO 11 I=1, M
XI(I, 1) = R(I)*COS(0.1745329*THEA(I))
XI(I, 2) = R(I)*SIN(0.1745329*THEA(I))
XI(I, 3) = Z(I)
11 CONTINUE

WRITE (6, 611)

611 FORMAT (/1X, 3F10.3, 1X, 10A8)

WRITE (6, 612) (XI(I), I=1, M)

612 FORMAT (3F10.3, 1X)

C CHECK FOR ALL Z'S EQUAL
DO 6 I=2, M
ZI = XI(I) - XI(1)
IF (ABS(ZI), LE, 1.0, **(-5)) GO TO 6
GO TO 5

6 CONTINUE
P(1) = 0.
P(2) = 0.
P(3) = 1.
PX = 0.
PY = 0.
PZ = 1.
GO TO 19

C CALCULATE XTRANSPOSE TIMES X
5 DO 12 K=1, 3
DO 12 L=1, 3
XX(K, L) = 0.
12 XX(K, L) = XX(K, L) + XI(K)*XI(L)

C CALCULATE XTRANSPOSE TIMES 1
DO 15 K=1, 3
XT(K, 1) = 0.
15 XT(K, 1) = XT(K, 1) + XI(K)

CALL MATINV(XX, 3, XT, 1, DET)

C CALCULATE VALUES OF Z IN PLANE
DO 720 I=1,M
720 X(I,3)=(1.-XT(1,1))*X(I,1)-XT(2,1)*X(I,2)/XT(3,1)
C CALCULATE UNIT VECTOR PERPENDICULAR TO PLANE
DEN=SQRT(X(I,1)**2+X(I,2)**2+X(I,3)**2)
DO 25 I=1,3
25 P(I)=XT(I,1)*XT(I,1)/DEN
PX=P(1)
PY=P(2)
PX=P(3)
19 TWIST=ACOS(PX)--57.31958
TILT=ACOS(PY)--57.2958
WRITE (6,300) TWIST,TILT,X(6,1)
300 FORMAT (/I1,6X,TWIST=*,F10.2,*,TILT=*,F10.2,*,D=*,F10.2)
READ (5,20) BASE,NO
20 FORMAT (A3,17)
READ (5,120) (TDIR(I),TMAG(I),I=1,NO)
120 FORMAT (8F10.2)
DO 22 J=1,NO
22 READ (5,21) (POLES(J,1),I=1,M)
21 FORMAT (8F10.4)
C CALCULATE UNIT VECTOR ALONG C5-C6 FOR PYRIMIDINE OR C4-C5 FOR PURINE
IF (BASE.EQ.3HPUR) GO TO 100
CALL UNITV(X,5,6,U)
GO TO 101
100 CALL UNITV(X,5,4,U)
101 CONTINUE
WRITE (6,613)
613 FORMAT (/I1,6X,Y,Z COORDINATES OF DIRECTION UNIT VECTORS FOR TRANS,
ITIONS AS LISTED BELOW*)
DO 26 I=1,NO
DIR=-TDIR(I)
WRITE (6,549) DIR,(U(I),I=1,3),(P(I),I=1,3)
549 FORMAT (10X,F8.3,F(2X,F10.6))
CALL BRTRAN(DIR,U,P,TX,TY,TZ)
WRITE (6,612) TX,TY,TZ
CALL PHITHE(TX,TY,TZ,A,B)
PHI(I)=A*57.2958
THE(I)=B*57.2958
26 CONTINUE
C FIND CENTERS OF TRANSITION MOMENTS
DO 31 J=1,NO
CALL CHARGE(POLES,M,J,Q,X,QCHEK)
IF (ABS(QCHEK).GE.10.*(-4)) GO TO 665
QX(J)=Q11
QY(J)=Q12
QZ(J)=Q13
31 CONTINUE
READ (5,401) NNO
401 FORMAT (I2)
READ (5,400) (BDIR(I),BMAG(I),I=1,NNO)
400 FORMAT (8F10.3)
C CALCULATE BACKGROUND POLARIZABILITIES
DO 41 I=1,NNC
DIR=-BDIR(I)
CALL BRTRAN(DIR,U,P,BX,BY,BZ)
WRITE (6,612) BX,BY,BZ
41 CONTINUE
IF (BASE, EQ, 3MPUR) GO TO 200
K=3
L=5
MM=1
DO 251 LL=7, M
251 BQ(I, LL)=0.
GO TO 202
200 K=6
L=7
MM=4
BQ(I, 1)=0.
BQ(I, 2)=0.
BQ(I, 3)=0.
DO 252 LL=10, M
252 BQ(I, LL)=0.
202 DO 201 JD=1, 2
A=X(K, I)-X(MM, I)
B=X(L, I)-X(MM, I)
C=-0.51*(BMAG(1)/4.802981)*AX
D=X(K, 2)-X(MM, 2)
E=X(L, 2)-X(MM, 2)
F=-0.51*(BMAG(1)/4.802981)*BY
DET=A*E-B*D
BQ(I, K)=(C*E-B*F)/DET
BQ(I, L)=(A*F-C*D)/DET
BQ(I, MM)=-BQ(I, K)-BQ(I, L)
IF (K, EQ, 6) GO TO 203
K=2
L=6
MM=6
GO TO 201
201 CONTINUE
CALL PHITHEIPX, IPY, IPZ, A, B
PHII(I)=A*57.2958
THE(I)=B*57.2958
41 CONTINUE
C FIND CENTERS OF BACKGROUND TRANSITION MOMENTS
DO 71 J=1, NNC
CALL CHARGE(BQ, M, J, Q, X, QCHEK)
IF (ABS(QCHEK), GE, 1.0E(-4)) GO TO 665
QBX(J)=Q(1)
QBY(J)=Q(2)
QRZ(J)=Q(3)
71 CONTINUE
READ (5, 81) PMAG
81 FORMAT (F10, 3)
AM=M
DO 80 I=1, 3
PERP(I)=0.
DO 80 J=1, M
80 PERP(I)=PERP(I)+X(I, J)/AM
CALL PHITHEIPX, PY, PZ, A, B
WRITE (6, 612) PX, PY, PZ
A = A * 57.2958
B = B * 57.2958
NNT = NNO + NNO
MNO = NNO
IF (NNT .GT. 9) MNC = 9 - NO
DO 631 I = 1, NO
QRR = (OX(I) * OX(I)) + (QY(I) * QY(I))
QR(I) = SQRT(QRR)
QQ = OX(I)
IF (ABS(QQ).LT. C001) GO TO 630
QT = QY(I) / OX(I)
QT(I) = ATAN(QT(I) * 57.2958)
IF (OX(I).LT. 0) QT(I) = QT(I) + 180.
GO TO 631
630 QT(I) = 90.
IF (QY(I).LT. 0) QT(I) = -90.
631 CONTINUE
DO 632 I = 1, NNO
QRR = (QBX(I) * QBX(I)) + (QBY(I) * QBY(I))
QB(I) = SQRT(QRR)
QQ = QBX(I)
IF (ABS(QQ).LT. C001) GO TO 633
QT = QBY(I) / QBX(I)
QB(I) = ATAN(QT(I) * 57.2958)
IF (QBX(I).LT. 0) QB(I) = QB(I) + 180.
GO TO 632
633 QB(I) = 90.
IF (QBY(I).LT. 0) QB(I) = -90.
632 CONTINUE
PPR = (PERP(I)**2) + (PERP(2)**2)
PERP = SQRT(PPR)
QQ = PERP(1)
IF (ABS(QQ).LT. C001) GO TO 634
PPT = PERP(2) / PERP(1)
PERPT = ATAN(PPT) * 57.2958
IF (QBX(I).LT. 0) PERPT = PERPT + 180.
GO TO 635
634 PERPT = 90.
IF (PERP(2).LT. 0) PERPT = -90.
635 CONTINUE
WRITE (6, 301)
301 FORMAT (/// *POSITIONS OF NUCLEI AND THEIR MONPOLIES*)
WRITE (6, 302)
DO 500 I = 1, M
500 WRITE (6, 303) (R(I), THETA(I), XI(I), Z(I), (POLES(J, I), J = 1, NO), (BQ(I), I = 1, NNO))
IF (NNT .LE. 9) GO TO 510
II = MNO + 1
DO 501 I = 1, M
501 WRITE (6, 303) (R(I), THETA(I), XI(I), Z(I), (BQ(I, I), I = 1, NNO)
510 CONTINUE
WRITE (7, 333) ((R(I), THETA(I), Z(I)), I = 1, M)
333 FORMAT (9F8.3)
DO 335 J = 1, NO
335 WRITE (7, 334) (POLES(J, I), I = 1, M)
DO 336 L=1,NNO
336 WRITE (7,334) (BQ(I,I),I=1,M)
334 FORMAT (8F10.6)
WRITE (6,304)
304 FORMAT (/* POSITIONS AND DIRECTIONS OF TRANSITIONS*/)
WRITE (6,305)
16X*,PHI*,5X*,MAGNITUDE*,3X*,DEVOE ANGLE*/
WRITE (6,306) (QXI(I),QYI(I),QRI(I),QT(I),QZI(I),THE(I),PHI(I),
ITMAG(I),TDIR(I),I=1,NO)
306 FORMAT (15,8F10.3,3X,Flu.3)
QA=0.
QB=0. 
QC=0.
DO 337 I=1,NG
337 WRITE (7,338) OR(I),QT(I),QZ(I),THE(I),PHI(I),QA,QB,QC,ITMAG(I)
338 FORMAT (9F8.3)
DO 306 I=1,NNO
NPD=NPD+I
3061 WRITE(6,3061) NPD,QBX(I),QBY(I),QBR(I),QBT(I),QBZ(I),THE(I),
1BPHI(I),BMMAG(I),BDIR(I)
DO 339 I=1,NNO
339 WRITE (7,339) QBR(I),QBT(I),QBZ(I),THE(I),BPHI(I),QA,QB,QC,
BMMAG(I)
L=NPD+NNO+1
WRITE (6,3061) L,PERP(I),PERP(2),PERPR,PERPT,PERP(3),B,A,PMAG
WRITE (7,338) PERPR,PERPT,PERP(3),B,A,QA,QB,QC,PMAG
C OUT OF PLANE Q CALCULATED FOR .75 A ABOVE AND BELOW PLANE OF BASE
QPERP=PMAG/(1.54*80298*M)
WRITE (6,640) QPERP
640 FORMAT (/*OUT OF PLANE Q ARE PLUS OR MINUS*,F10.6,* FOR POINTS
175 A FROM PLANE OF BASE*/)
DO 340 I=1,M
IQ=2*I
QPP(I,I)=QPERP
QPP(I-1,I)=(-QPERP)
340 CONTINUE
WRITE (7,334) (QPP(I),I=1,M)
GO TO 6651
665 WRITE(6,307) J
307 FORMAT (* SUM OF MONPOLES FOR TRANSITION*12,* DOES NOT = ZERO*)
GO TO 666
6651 READ (5,308) NOGO
308 FORMAT (II)
IF (NOGO) 600,666,600
666 CONTINUE
END
SUBROUTINE PHITHEIC1,C2,C3,P,T)
  T=ACOS(C3)
  IF (ABS(T) .LE. 10.**(-6)) GO TO 3
  P=ACOS(C1/SIN(T))
  IF (C2) 2,3,4
  2 P=-P
  GO TO 4
  3 P=0.
  4 RETURN
END

SUBROUTINE UNITY(X,N1,N2,U)
  DIMENSION X(20,3),U(3)
  DEN=SQRT((X(N1,1)-X(N2,1)**2+(X(N1,2)-X(N2,2)**2+(X(N1,3)-X(N2
  1,3)**2))
  DO 25 I=1,3
  25 U(I)=(X(N1,1)-X(N2,1)))/DEN
  RETURN
END

SUBROUTINE BTRAN(DIP,U,P,TX,TY,TZ)
  DIMENSION U(3),P(3)
  ST=SIN(.01745329*R))
  CT=COS(.01745329*R))
  TX=ST*U(2)*P(3)-U(3)*P(2)+U(1)*CT
  TY=ST*U(3)*P(1)-U(1)*P(3)+U(2)*CT
  T7=ST*U(1)*P(2)-U(2)*P(1)+U(3)*CT
  RETURN
END

SUBROUTINE CHARGE(A,M,J,Q,X,QCHEK)
  DIMENSION A(10,20),Q(3),X(20,3)
  DO 10 N=1,3
  10 Q(N)=0.
  QT=0.
  QCHEK=0.
  DO 9 I=1,M
    QT=QT+ABS(A(J,I))
    QCHEK=QCHEK+A(J,I)
  9 DO 8 N=1,3
  8 Q(N)=Q(N)+ABS(A(J,I))*X(I,N)
  DO 8 N=1,3
  8 Q(N)=Q(N)/QT
  RETURN
END
SUBROUTINE MATINV(A,N,B,M,DETERM)
DIMENSION IPIVOT(3),A(3,3),B(3,1),INDEX(3,2),PIVOT(3)
EQUIVALENCE (IROW,JROW), (ICOLUM,JCOLUM), (AMAX, T, SWAP)

10 DETERM=1.0
15 DO 20 J=1,N
20 IPIVOT(J)=0
30 DO 550 I=1,N
40 AMAX=0.0
45 DO 105 J=1,N
50 IF (IPIVOT(J)-1) 60, 105, 60
60 DO 100 K=1,N
70 IF (IPIVOT(K)-1) 80, 100, 740
80 IF (ABS(A(I,J))-ABS(A(J,K))) 85, 100, 100
85 IROW=J
90 JCOLUMN=K
95 AMAX=A(J,K)
100 CONTINUE
105 CONTINUE

IF (AMAX) 110, 800, 110
110 IPIVOT(JCOLUMN)=IPIVOT(JCOLUMN)+1
130 IF (IROW-JCOLUMN) 140, 260, 140
140 DETERM=-DETERM
150 DO 200 L=1,N
160 SWAP=A(IROW,L)
170 A(IROW,L)=A(JCOLUMN,L)
180 A(JCOLUMN,L)=SWAP
205 IF (M) 260, 260, 210
210 DO 250 L=1,M
220 SWAP=B(IROW,L)
230 B(IROW,L)=B(JCOLUMN,L)
250 B(JCOLUMN,L)=SWAP
260 INDEX(1,1)=IROW
270 INDEX(1,2)=ICOLUMN
310 PIVOT(1)=A(JCOLUMN,ICOLUMN)
320 DETERM=DETERM*PIVOT(1)
330 A(JCOLUMN,ICOLUMN)=1.0
340 DO 350 L=1,N
350 A(JCOLUMN,L)=A(JCOLUMN,L)/PIVOT(1)
355 IF (M) 380, 380, 360
360 DO 370 L=1,M
370 B(JCOLUMN,L)=B(JCOLUMN,L)/PIVOT(1)
380 DO 550 L=1,N
390 IF (L-JCOLUMN) 400, 550, 400
400 T=A(L,JCOLUMN)
420 A(L,JCOLUMN)=0.0
430 DO 450 L=1,N
450 A(L,L)=A(L,L)-A(JCOLUMN,L)*T
455 IF (M) 550, 550, 460
460 DO 500 L=1,M
500 B(L,L)=B(L,L)-B(JCOLUMN,L)*T
550 CONTINUE
600 DO 710 L=1,N
610 L=N+1-I
620 IF (INDEX(L,1)-INDEX(L,2)) 630, 710, 630
630 JROW=INDEX(L,1)
640 JCOLUMN=INDEX(L,2)
650 DO 705 K=1,N
660 SWAP=A(K,JROW)
670 A(K,JROW)=A(K,JCOLUMN)
700 A(K,JCOLUMN)=SWAP
705 CONTINUE
710 CONTINUE
740 RETURN
800 DETERM = 0.
RETURN
END
V. **Program CDORD** — Revised version of program by M.S. Itzkowitz. Allows Kronig-Kramers transform to be performed on data as a function of wavelength, then interpolates output to give data as a function of frequency.

**Revised DATA CARDS:**

1. **Title card** Format (12A6) Statements on these cards are printed at beginning of output.

2. 

3. **TYPE, LMIN, LCDMAX, DELTAL, LMAX, LPLOT, LMPAX**

   Format (A6, 4X, 6F10.6) Note: LMIN through LMAX are designated by program to be real variables, not integer.

   **TYPE** = "CD-ORD" if input is in form of absorption (CD, Im \( \alpha \)) "ORD-CD" if input is in form of refraction (ORD, Re \( \alpha \))

   **LMIN** = lowest wavelength (in nm) desired in transform

   **LCDMAX** = highest wavelength for which data to be transformed is read in on cards

   **DELTAL** = wavelength increment on data cards

   **LPMAX** = highest wavelength desired on output plot

   **LPLOT** = output plot wavelength increment

   **LMAX** = highest wavelength desired in transform

   Minimum wavelength in output plot is determined by \( \text{LPMAX} - 100(\text{LPLOT}) \), i.e. to give 101 points.

   **LCDMAX may be equal to LMAX.** They are not equal when absorption-type data is read in on cards and the absorption is zero at long wavelengths. Then data is punched on cards from LCDMAX to LMIN, and the program automatically assigns the value zero to the absorption from \( (\text{LCDMAX} + \text{DELTAL}) \) to LMAX

4. **INPLT, IOPHT, IBACK, IBPLT, IFREQ** Formatted (5I1)

   If set equal to 1, the following occurs:
IIPLT  input data is plotted
IOPLT  transform is plotted
IBACK  backtransform is calculated
       (a check - should equal input
       in shape)
IBPLT  backtransform is plotted
IFREQ  transform versus wavelength is
       interpolated to give transform
       versus frequency

5. WMAX, WMIN, DELTAF  Note: Use blank
Format (6X,3F10.6) card here if IFREQ
above is set equal
to zero

WMAX = maximum frequency in kK (10^3 cm^-1)
for interpolation
WMIN = minimum frequency in kK (10^3 cm^-1)
for interpolation
DELTAF = frequency increment in kK (10^3
       cm^-1) for interpolation

NOTE: WMAX must be < \frac{1}{LMIN}

WMIN must be > \frac{1}{LMAX}

6*. Data to be transformed - Format (8F10.6) - starting
with LCDMAX value and ending with LMIN value in
steps of DELTAL. 500 points maximum.

7. Additional sets of data to be transformed may be
entered with series of cards 1 through 6.

8. END series of data sets with card reading $EOF in
columns 1-4
(Then End of Job card)

*Program is currently set up to accept input as ε vs. λ
(ε = extinction coefficient in 1/mole cm) and convert
it to output in the form of Re α and Im α (real and
imaginary polarizability in Å^3)

For other conversion, only 2 statements need be
changed. They currently read

ORD(I) = ORD(I) * WAVELT(I) * .000014525
CD(I) = CD(I) * WAVELT(I) * .000014525
PROGRAM CDORD(INPUT,OUTPUT,PUNCH)

C
C MAIN PROGRAM FOR CD-ORD TRANSFORMS (EITHER WAY)
C M.S. ITZKOWITZ 3/28/68
C
DIMENSION CD1500I,ORD1500I,FUNCT1500I
DIMENSION COMNT124I,WAVELT1500I
DIMENSION WNUM1500I,FREQ1500I,FORD1500I,FORD1500I
DATA JINPUT/6H INPUT/,JTRANS/6H TRANS/,JBACK/5H BACK/
DATA JORD/6H ORD/,JCD/6H CD/
DATA END/FIL16H$EOF/
DATA RP1/0.3183098861838/
REAL LMIN,LMAX,LPLOT,LMAX,LBAR,LBARSQ,LCMAX

901 FORMAT(12A6)
902 FORMAT(A6,4X,7F10.6)
903 FORMAT(8F10.6)
904 FORMAT(1H1,12A6,/1X,12A6,5X,A6,1X,A6)
905 FORMAT(10HHOWAVELNGTH,5X,2HCD,13X,3HORD,/(1XF10.2,2XE13.6,2XE13.6))
906 FORMAT(10I11)
907 FORMAT(10HHOWAVELNGTH,5X,3HORD,12X,2HCD,/(1XF10.2,2XE13.6,2XE13.6))
908 FORMAT(134HOBEST FIT OF DUROE TO ORD TAIL, A=E15.8,5X,5HLBAR=E15.8)
909 FORMAT(1H1,12A6,/1X,12A6)
405 FORMAT(10HHOWAVENMBER,5X,2HCD,13X,3HORD,/(1XF10.2,2XE13.6,2XE13.6))
406 FORMAT(10F8.2)
407 FORMAT(6X,3F10.6)

1 READ 901, (COMNT1),I=1,24
IF(COMNT1.EQ.ENDFIL) CALL EXIT
READ 902, TYPE,LMIN,LCMAX,DELTAL,LPLOT,LMAX
READ 906, JINPUT, JOUTPUT, JBACK, JFREQ
- READ 407, WMAX, WMIN, DELTA F
NPTS=((LMAX-LMIN)/DELTAL) + 1.01
DO 20 I=1,NPTS
WAVELT(I)=LMIN+FLOAT(I-1)*DELTA F
CONTINUE
K=(LMAX-LCMAX)/DELTAL + .01
M=K + 1.
DO 600 I=1,K
FUNCT(I)=0.
600 CONTINUE
READ 903, (FUNCT(I),I=M,NPTS)
INDEX=JINPUT
IF(TYPE.EQ.6HORD-CD) GO TO 100
IF(TYPE.EQ.6HCD-ORD) GO TO 200
PRINT 9901, TYPE
9901 FORMAT(1H1),A6,17H TYPE JOB UNKNOWN )
CALL EXIT
100 CONTINUE
C
C TRANSFORM FROM ORD TO CD
C
DO 105 I=1,NPTS
ORD(I)=FUNCT(NPTS-I+1)
ORD(I)=ORD(I)*WAVELT(I)*.014525
FIT TAIL #N# POINTS TO A SINGLE DRUDE TERM

N# = 4
A = 0.
B = 0.
C = 0.
D = 0.
XN = N#
DO 110 I = 1, N#
INDEX = NPTS - I + 1
Y = WAVELT(INDEX)**2
IF(ORD(INDEX),EQ,0.0) GO TO 115
X = Y/ORD(INDEX)
A = A + X*Y
B = B + Y
C = C + Y**2
D = D + X
110 CONTINUE
DET = XN*C - B**2
BETA = (D*C - A*B)/DET
IF(BETA,EQ,0.0) GO TO 115
ALPHA = (B*D - XN*A)/DET
A = -1./BETA
LBARSQ = BETA/ALPHA
IF(LBARSQ,EQ,0.0) GO TO 115.
LBAR = SQRT(LBARSQ)
IF(LBAR .GT. LMAX) GO TO 115
GO TO 116.
115 CONTINUE
A = 0.
LBAR = 0.
116 CONTINUE
PRINT 909, COMNT
PRINT 908, A, LBAR

BEST DRUDE = A*LSQ*LBARSQ/(LSQ-LBARSQ)

WAVCUT = LMAX + 0.5*DELTAL
DO 120 I = 1, NPTS
FUNCT(I) = ORD(I)/WAVELT(I)
120 CONTINUE
DO 125 I = 1, NPTS
CALL HILBRT(NPTS, FUNCT, LMIN, LMAX, WAVELT(I), DELTAL, HPLUS, HMINUS)
TAIL = 0.
IF(A,EQ,0.0) GO TO 121
WVDENV = WAVELT(I)**2 - LBARSQ
IF(WVDENV,EQ,0.0) GO TO 121
TAIL = RPI*A*WAVELT(I)**2 - LBARSQ/WVDENV
TAil=TAIL*(ALOG((WAVELT(I)+WAVCUT)/(WAVCUT-WAVELT(I))))-LBAR/
$WAVELT(I)+ALOG((LBAR+WAVCUT)/(WAVCUT-LBAR))$

121 CONTINUE
CD(I)=WAVELT(I)*RPI*(HPLUS-HMINUS)+ TAIL
125 CONTINUE
PRINT 937,(WAVELT(I),ORD(I),CD(I),I=1,NPTS)
IF(1OPLT.EQ.0) GO TO 131
JTYPE=JCD
PRINT 904,COMNT,JTYPE,JINDEX
CALL PRNPLT(WAVELT,CD,LPMAX,LPLOT,XMAX,XINCR,0,1,NPTS)
131 IF(IBACK.EQ.0) GO TO 1
IBACK=0
IOPLT=IOPLT
GO TO 201
C
200 CONTINUE
C TO ORD TRANSFORMATION
C
DO 205 I=1,NPTS
CD(I)=FUNCTIONPTS-I+1)
CD(I)=CD(I)*WAVELT(I)*.014525
205 CONTINUE
IF(IOPLT.EQ.0) GO TO 300
JTYPE=JCD
PRINT 904,COMNT,JTYPE,JINDEX
CALL PRNPLT(WAVELT,CD,LPMAX,LPLOT,XMAX,XINCR,0,1,NPTS)
300 JINDEX=JTRANS
201 CONTINUE
DO 210 I=1,NPTS
FUNCTION=CD(I)/WAVELT(I)
210 CONTINUE
DO 220 I=1,NPTS
CALL HILBRTNPTS,FUNCTIONPTS,FUNCTIONPTS,LMIN,LMAX,WAVELT(I),DELTAL,HPLUS,HMINUS)
ORD(I)=WAVELT(I)*RPI*(HPLUS+HMINUS)
220 CONTINUE
PRINT 909,COMNT
PRINT 905,(WAVELT(I),CD(I),ORD(I),I=1,NPTS)
IF(IOPLT.EQ.0) GO TO 700
JTYPE=JORD
PRINT 904,COMNT,JTYPE,JINDEX
CALL PRNPLT(WAVELT,ORD,O,LPLOT,XMAX,XINCR,0,1,NPTS)
700 IF(IFREQ.EQ.0) GO TO 221
400 DO 401 I=1,NPTS
WNUM(I)=10000./WAVELT(I)
401 CONTINUE
MPTS=(WMAX-WMIN)/DELTAF + 1.01
DO 402 J=1,MPTS
FREQ(J)=WMAX-FLOAT(J-1)*DELTAF
403 IF(FREQ(J).LT.WNUM(I+1)) I=I+1
IF(FREQ(J).LT.WNUM(I+1)) GO TO 403
FCDIJ=CD(I+1)*(FREQ(J)-WNUM(I+1))/(WNUM(I)-WNUM(I+1))$CD(I+1)
ORD(I)=ORD(I)-ORD(I+1)*(FREQ(J)-WNUM(I+1))/(WNUM(I)-WNUM(I+1))$ORD(I+1)
402 CONTINUE
PRINT 405, (FREQ(J), FCD(J), FORD(J), J=1, MPTS)
PUNCH 406, (FCD(J), J=1, MPTS)
PUNCH 406, (FORD(J), J=1, MPTS)
221 IF (IBACK.EQ.0) GO TO 1
  JINDEX=JBACK
  IBACK=0
  IOPLT=18PLT
  GO TO 101
END
SUBROUTINE HILBERT(NPTS,FUNCT,LMIN,LMAX,LPRIME,DELTAL,HPLUS,HMINUS)
  HILBERT TRANSFORM PROGRAM
  M.S. TIKZOMITZ 4/16/68

  INTEGRALS FROM LMIN TO LMAX OF
    (FUNCT(L)/LPRIME + L) = HPLUS
  AND
    (FUNCT(L)/LPRIME - L) = HMINUS

  NPTS POINTS STARTING AT LMIN AND ENDING AT LMAX IN INTERVALS OF
  DELTAL ARE USED FOR THE INTEGRAL. TRAPEZOIDAL RULE IS USED FOR INTEGRATIONS
  FOR HPLUS FOR THE ENTIRE RANGE AND FOR HMINUS IN THE TWO
  RANGES LMIN TO LPRIME-DELTAL AND LPRIME+DELTAL TO LMAX. EXPANSION
  UP TO THE THIRD DERIVATIVE OF FUNCT(L) IS USED FOR THE INTERVAL FROM
  LPRIIME-DELTAL TO LPRIME + DELTAL. NUMERICAL EVALUATION OF THE
  DERIVATIVES USES THE FIVE POINT FORMULA FROM ABRAMOWITZ AND STEGUN

  ASSUMES THAT FUNCT(1). CONTAINS FUNCT AT LMIN AND FUNCT(NPTS)
  CONTAINS FUNCT AT LMAX

  DIMENSION FUNCT(NPTS)
  REAL LMAX,LMIN,LPRIME
  HMJNUS=0.
  HPLUS=0.
  LPRIME=1.*((LPRIME-LMIN)/DELTAL)
  LMIN=LMIN/DELTAL-1.
  DO 100 I=1,NPTS
    HPLUS=HPLUS+FUNCT(I)/FLOAT(I*LMIN+1*LPRIME)
    IF((LPRIME.LT.1.OR.IPRIME.GT.NPTS)) GO TO 150
    IF((LPRIME.EQ.2)) HMJNUS=HMJNUS+0.1111111*FUNCT(IPRIME-2)
    IF((LPRIME.GT.1)) HMJNUS=HMJNUS+0.7777778*FUNCT(IPRIME-1)
    IF((LPRIME.LT.NPTS)) HMJNUS=HMJNUS+0.7777777*FUNCT(IPRIME+1)
    IF((LPRIME.LT.NPTS-1)) HMJNUS=HMJNUS+0.1111111*FUNCT(IPRIME+2)
  CONTINUE

  RETURN
END

SUBROUTINES PRNPLT AND PISCAL as in ROTOL1
VI. PROGRAM FINALE - from thesis of M. Warshaw, modified to calculate $f$, $|\psi_1|$, and $|\psi_1|^2$ from extinction coefficient as a function of wavelength.

```
PROGRAM FINALE (INPUT, OUTPUT)
C NUMERICAL INTEGRATION OF ABSORPTION CURVES USING SIMPSON'S RULE
C, PROGRAM SET UP FOR INPUT IN FORM OF EXT.XLAMBD.A - IF INPUT IS
C EXT. ONLY, REMOVE CARDS NUMBERED 1 AND 2
DIMENSION FLA(100), OD(100), FNU(100), TITLE(14), EXT(100), DATA(100)
3 READ 10, ITITLE(I), I=1,14
10 FORMAT (14A4)
5 READ 11, START1, START2, STOP, DELTA, CONV, LRVRSE
11 FORMAT (4F10.1, F10.0, I1)
IF (START1 .LT. 6,6,50)
50 M=(STOP-START1)/DELTA + 1.
READ 12, (DATA(I), I=1, M)
12 FORMAT (8F10.2)
IF (LRVRSE.EQ.0) GO TO 60
DO 70 I=1, M
70 CONTINUE
GO TO 80
60 CONTINUE
DO 90 I=1, M
90 CONTINUE
80 CONTINUE
FLA(I)=START1
FNU(I)=10000000./FLA(I)
EXT(I)=OD(I)*CONV/1000.
DO 21 I=1, M
21 CONTINUE
DO 22 I=1, N
22 CONTINUE
L=(START2-START1)/DELTA + 1.
G=CD(I)/FLA(I)**2
D=OD(I)/FLA(I)
M=(M-L)/2
DO 23 I=1, N
23 CONTINUE
```

DIPSTR = DELTA * CONV * (9.186 * 1.31 * 10^(-6))
TRANMO = (DIPSTR) ** 5

PRINT 13
13 FORMAT (1H1, 45X, 43H NUMERICAL INTEGRATION OF ABSORPTION CURVES /)
PRINT 69, START2

PRINT 69, START2
69 FORMAT (100X, 18H INTEG. CUT-OFF = F6.1)
PRINT 14, (TITLE(I), I=1, 14)

PRINT 15
15 FORMAT (30X, 14A6 //)

PRINT 16
16 FORMAT (1X, 4133H LAMBDA C.D. FREQ. )//)
PRINT 17, (FLA(I), DD(I), FNU(I), I=1, M)
17 FORMAT (4(F13.1, F10.3, F10.0)//)

PRINT 18
18 FORMAT (/39X, 46H * * * * * * * * * * * * * * * * //)

PRINT 19
19 FORMAT (/50X, 22H INTEGRALS AND MOMENTS //)
PRINT 91, (OSC, DIPSTR, TRANMO)

PRINT 91, (OSC, DIPSTR, TRANMO)
91 FORMAT (8X, 23H OSCILLATOR STRENGTH = F6.4, 24H DIPOLE STRENGTH = F6.3, 35H DEBYE**2 TRANSITION MOMENT = F6.3, 6H DEBYE// I)
PRINT 13
PRINT 14, (TITLE (I), I=1, 14)

PRINT 96
96 FORMAT (53X, 25H EXTINCTION COEFFICIENTS //)
PRINT 97

PRINT 97
97 FORMAT (1X, 4129H LAMBDA EXTINCTI)
PRINT 98

PRINT 98
98 FORMAT (1X, 4129H XE-3 //)
PRINT 99, (FLA(I), EXT(I), I=1, M)

PRINT 99, (FLA(I), EXT(I), I=1, M)

GO TO 3
6 STCP
END
This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Energy Research and Development Administration, 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.