UCRL-17966
cy. 2

# University of California

# Ernest O. Lawrence Radiation Laboratory

TEMPERATURE DEPENDENT PROPERTIES
OF DINUCLEOSIDE PHOSPHATES

Robert Charles Davis
(Ph. D. Thesis)

November 1967

RECEIVED
LAWRENCE
RADIATION LABORATORY

LIBRARY
DOCUMENTS

TWO-WEEK LOAN COPY

*This is a Library Circulating Copy which may be borrowed for two weeks. For a personal retention copy, call Tech. Info. Division, Ext. 5545*

UCRL-17966
cy. 2

## 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.

UCRL-17966

UNIVERSITY OF CALIFORNIA
Lawrence Radiation Laboratory
Berkeley, California

AEC Contract No. W-7405-eng-48

# TEMPERATURE DEPENDENT PROPERTIES OF DINUCLEOSIDE PHOSPHATES

Robert Charles Davis
(Ph. D. Thesis)

November 1967

# TABLE OF CONTENTS


ABSTRACT

INTRODUCTION 1

MEASUREMENT OF TEMPERATURE DEPENDENT PROPERTIES OF DINUCLEOSIDE PHOSPHATES 3

Review of Past Experimental Work 3

Physical Origins of the Properties to be Measured 14

Optical Rotatory Dispersion 17

Ultraviolet Absorption 19

Nuclear Magnetic Resonance 23

Models 25

Two-State Model 26

Multi-State Model 27

Oscillating Dimer Model 28

Generalized Energy Surfaces 32

Materials and Methods 33

Nucleic Acid Material 33

Non-Biological Materials 34

Determination of Concentrations 36

Instruments and Measurements 37

Results and Discussion 43

Two-State Model 43

Oscillating Dimer Model 64

Low Temperature Measurements 67

Nuclear Magnetic Resonance 77

Comparisons among Dinucleoside Phosphates 81

Comparison with the Literature 91

Energetic Considerations 98

CALCULATIONS OF THE STABILITIES OF NUCLEIC ACID STRUCTURES 115

Introduction 115

Methods 128

Results and Discussion 156

Nucleosides 156

Internucleoside Bonds 178

CpC 194

Group I 203

Groups II and IV 206

Groups III, V, VI, VII and IX 206

Group VIII 207

Paths between Conformations 223

Entropy 231

Conclusion 237

ACKNOWLEDGEMENTS 241

REFERENCES 243

APPENDIX A - ABBREVIATIONS USED 258

APPENDIX B - COMPUTER PROGRAMS 259

## ABSTRACT

In order to learn about single-strand nearest neighbor interactions in ribonucleic acids the temperature dependent properties of dinucleoside phosphates have been studied. Optical rotatory dispersion, ultraviolet absorption and proton magnetic resonance were used to probe the thermal disordering of these compounds. Two models were used to analyze the data. For the two-state model, a molecule was considered to be in either the ordered or disordered form with no permissible conformations between them. For the oscillating dimer model, the two bases of a dinucleoside phosphate were constrained by a single torsional harmonic potential. The properties of dinucleoside phosphates in dilute salt solutions (from 0°C to 90°C) are compatible with either model. In the former case $\Delta H^{o}$ for disordering ranges from 4.8 to 8.4 kcal per mole while in the latter case the force constant ranges from 120 to 340 cal per mole radian$^2$. This force constant means that approximately 2 kcal per mole are required for oscillation of ±180 degrees around the equilibrium position.

Measurements in concentrated LiCl solutions (where the range of experimentally accessible temperatures is almost doubled) show that neither of these models is adequate. The thermal disordering process is one in which the bases can assume a number of intermediate conformations. Increasing temperatures cause motion of the bases away from one another. Proton magnetic resonance and ultraviolet absorption measurements show that there are appreciable interactions between the bases at high temperatures, suggesting that

disordering involves motions of the bases parallel to one another, without appreciable solvent ever coming between them.

A wide variation of tendencies to "stack" is observed among different dinucleoside phosphates. The midpoint of the disordering transition (as determined by optical rotatory dispersion) is approximately + 5°C for "stacked" dinucleoside phosphates such as UpU and UpC. This suggests that sharp turns in a polynucleotide structure are more likely to occur in regions rich in uridine, while single-strand regions rich in adenosine will probably resemble worm-like coils.

Calculations of the potential energies of conformations of nucleic acid fragments were also carried out. Molecules were considered to consist of rigid sections (e.g. a base) connected by single covalent bonds about which rotation can take place. Interatomic potentials included charge-charge, charge- polarizability, polarizability-polarizability and steric repulsion terms. These calculations indicate that pyrimidine nucleosides are all in the anti conformation while purine nucleosides may be either syn or anti. Energetically stable torsional angles calculated for larger systems generally agree with those found in polynucleotide fiber structures. Some possible origins of the differences between polynucleotides with ribose and deoxyribose sugars are found and discussed. Calculations on CpC indicate that the most stable conformations are compact structures, with close contacts between bases and sugars and relatively little base-base interactions. This is probably caused by not including solvent effects. Methods of correcting inadequacies and extending the calculations to larger and more interesting systems are discussed.

# INTRODUCTION

One of the most important contributions a physical chemist can make to biology is the elucidation of the structure of biologically important macromolecules. The three dimensional conformation of the biopolymer is an invaluable help in determining the biologically important role the molecule plays as well as the mechanism of action. The mechanism of DNA as the genetic material[214,215] and the enzymatic mechanism of lysozyme[142,164] have been clarified after their three dimensional structures were elucidated.

The class of biopolymers to which this work is applied is the ribonucleic acids. Although much is already known about their importance in the translation of the genetic message into the amino acid sequence of proteins, the mechanisms of these actions remain for the most part unclear[213]. Knowledge of the secondary and tertiary structures of RNAs would be invaluable in elucidating their biological roles. Most RNAs lack the high degree of symmetry shown by DNAs, making the study of their three dimensional structure considerably more difficult. It is to our advantage to divide this large problem into a number of smaller ones. The specific goal of this work is to understand the interaction between neighboring nucleotides, covalently bonded to one another by a single phosphodiester linkage. In order to separate this interaction from others, dinucleoside phosphates are studied. When dinucleoside phosphates are in sufficiently dilute solution so that intermolecular

interactions between solute molecules can be neglected, the only interactions remaining are those between solute and solvent and those between the various parts of the dinucleoside phosphate itself, the purine and pyrimidine bases, the ribose sugars and the phosphate group. These interactions will also occur in RNA and will be especially important in single-strand regions where the only nearest neighbors of a nucleotide are nucleotides which are probably oriented in a manner very similar to that in a dinucleoside phosphate[29,30].

In this work an attempt is made to determine the conformations of dinucleoside phosphates in aqueous solution as a function of temperature. In addition, the variations in conformation resulting from differing sequences of nucleosides are considered. These conformations as well as their sequence dependence can then be applied to RNAs, especially single-strand RNAs.

The temperature dependence of the conformations of dinucleoside phosphates gives information about the energy (both enthalpic and entropic) of these conformations. Knowledge of the nature of the forces which are important in the case of dinucleoside phosphates would be valuable in understanding the stabilities of single-strand as well as other RNA structures. There are forces in multiple-strand regions of RNAs which are not present in dinucleoside phosphates (e.g. base-base hydrogen-bonding). Even if these forces are largest, there will certainly be important contributions to the structure of all RNAs by the forces dominant in dinucleoside phosphates.

# MEASUREMENT OF TEMPERATURE DEPENDENT PROPERTIES OF DINUCLEOSIDE PHOSPHATES

## REVIEW OF PAST EXPERIMENTAL WORK

A great deal of experimental work has already been completed which bears on the question of the conformation of dinucleoside phosphates. Although this evidence ranges from the direct to the circumstantial, all of it is relevant and is helpful in formulating our approach to this problem.

Probably the most convincing data come from X-ray diffraction studies. A great many studies have been carried out on fibers of multiple-strand RNA and DNA polymers[42]. Unfortunately no work has been reported on fibers of single-strand polynucleotides. The major problem obstructing this work is the lack of success in pulling fibers of these single-strand polynucleotides from solutions. In the cases of the polynucleotides which have been studied, the bases of neighboring nucleotides are found parallel to one another, approximately 3.4Å from one another. The angle between the plane of the base and the screw axis of the helix ranges from 70° to 90°. All helices found have been right-handed with 8 to 12 residues per turn.

The X-ray diffraction studies of fibers was the first work suggesting a "stacked" conformation of a dinucleoside phosphate, where the bases were parallel to one another and more or less the minimum distance apart, in an orientation similar to that found in two adjacent nucleotides of a polynucleotide chain.

There has been only one completed X-ray diffraction study of a single crystal of a nucleic acid material more complicated than a nucleotide[172]. The structure of adenylyl (2'-5') uridine (a compound not normally found in nature) shows a "stacked" conformation. No complementary base-pairing was found although this compound does have the possibility of forming hydrogen-bonded base-pairs. These results again suggest a "stacked" orientation of bases of a dinucleoside phosphate in solution.

In spite of the fact that these studies have yielded much valuable information about the conformations of dinucleoside phosphates in solution they are not the ideal tools for our study. They measure the structure of molecules in single crystals or polycrystalline fibers. We are interested in the structures of molecules in aqueous solution. In crystals these molecules interact strongly with their neighbors. This is exemplified by the extensive intermolecular hydrogen-bonding found. Another disadvantage is that we would like to measure the conformations (or conformation dependent properties) of the molecules under continuously varying conditions (e.g. temperature, pH, solvent composition, etc.). In the case of crystal studies there is usually no variation possible. The environment of the fibers (i.e. temperature, humidity) can be varied in a continuous manner. Unfortunately, the changes found in polynucleotide fibers are usually highly cooperative ones. DNA, for example, exists in either the A, B, or C forms,

all distinct and quite different from one another. There are no observable intermediate forms.

Although results of X-ray diffraction studies have been invaluable in helping us to understand certain aspects of nucleic acid structure, we must turn to other techniques to study the conformations of dinucleoside phosphates.

Another source of information which is easily interpreted is the study of the colligative properties of bases and nucleosides. Measurements on aqueous solutions of purine and pyrimidine bases show marked decreases in the osmotic coefficients of the bases with increasing concentration in the range of 0.1 to 1 molar[21,176,194,196]. This is most easily explained in terms of aggregation, where the aggregate sizes may become quite large, exceeding ten monomer units. Similar studies, using ultracentrifugal techniques to measure molecular weights, reach similar conclusions[174, 204]. The interactions which result in aggregation are very dependent on the nature of the base, with purine bases interacting more strongly than pyrimidine bases. This is an important fact, suggesting that there may be different conformations of dinucleoside phosphates which depend on their base composition.

It is tempting to say that bases are aggregating in vertical "stacks", just as pancakes can be stacked on a breakfast plate. However the geometry of the aggregate formed can not be determined by thermodynamic measurements alone. (We will see later that nuclear magnetic resonance studies give exactly this information.) It is this lack of

structural information combined with the fact that they measure intermolecular rather than intramolecular interactions which make these techniques not particularly useful to us in the studies of dinucleoside phosphate conformations.

Quite similar to these are the measurements of the cooperativity of the helix-coil transition in multi-strand polynucleotides and the binding of small oligonucleotides to their complementary polynucleotides[6,81,113,114]. Much of the cooperativity of these helix-coil transitions can be explained on the basis of energetically favorable "stacking" of bases[40]. The first nucleoside binding to a complementary polynucleotide gains only the energy of base-pair formation. The next nucleoside gains not only this base-pairing energy, but also energy from the "stacking" interaction with the first nucleoside. Because it is easier to bind a nucleoside next to another nucleoside rather than next to unoccupied sites, the binding of these smaller complementary molecules to polynucleotides is cooperative.

Unfortunately these techniques have drawbacks similar to the studies of colligative properties. They give no direct information about conformations, one of the key points we are trying to determine. They also measure a multiple-strand interaction, something not of immediate interest to us.

There is other less direct evidence on the problem of the relative orientation of adjacent nucleotides of a single-strand oligonucleotide or polynucleotide. The dependence of the photodimerization of thymine on the

solvent has been studied[207]. Data from these experiments show that the quantum yield increases in solvents which are considered to increase "stacking" of bases. This is very reasonable because proximity of two bases enhances the rate of their photodimerization.

Studies of pyrimidine photodimerization are only of limited use to us. They can not be applied to all dinucleoside phosphates. In addition, quantitative interpretation of the data would be very difficult. The chemical mechanism of the dimerization is very complicated with a number of excited intermediates already found[22]. Effects of the environment (e.g. interaction with the neighboring base or interaction with the solvent which may also depend on the position of the neighboring base) are almost impossible to separate.

Since we are interested in the forces responsible for the structures of dinucleoside phosphates direct measurements for the thermodynamic parameters for transitions from one conformation to another would be very useful. Unfortunately, quantitative measurements of the energetic parameters of the "unstacking" process are sparse. Calorimetric measurements of the enthalpy and entropy of "stacking" of some nucleic acid bases and base analogues have been made[66,179]. In addition there are calorimetric measurements of the enthalpy of double and triple-strand formations at varying temperatures[24,163]. If one considers the temperature dependence of $\Delta H$ for the poly A + poly U

= poly (A+U) reaction to come from the change in "stacking" of bases in the single-strand polynucleotides with changing temperature, then $\Delta H$ for the "stacking" reaction can be determined. In spite of the fact that a great many assumptions have to be made in this analysis it does serve as a useful check for some of our later work.

Direct calorimetric measurements of the "stacked" to "unstacked" reaction of bases and nucleosides have many drawbacks. Because the geometries may be quite different, application of the results obtained to the problem of the structure of dinucleoside phosphates is difficult. In spite of the difficulties in direct calorimetric measurements, unambiguous values of the changes of enthalpy of dinucleoside phosphates under varying conditions would be invaluable. The way this could be determined would be to measure the partial molar heat capacities of dinucleoside phosphates in solution, then integrate over the temperature range of the ordered-disordered transition. Such measurements have been made for the helix-coil transition of DNAs and polypeptides[1,149]. Unfortunately, the transition region for the disordering of the bases of a dinucleoside phosphate is very broad compared to the sharp transitions of polynucleotides and polypeptides, making the experiment too difficult to do with sufficient precision at this time. One similar set of experiments has been reported for single-strand poly A. The heat of solution of poly A was measured calorimetrically at different temperatures. Using the

results of optical measurements on poly A to determine the change in "stacking" over this temperature range, a value of $\Delta H^o$ ("stacking") was calculated[56]. The dependence of this technique on optical measurements really places it with other optical methods in its usefulness and limitations. What we will say about optical studies will apply to a large extent to this type of measurement as well.

Up to this point we have not discussed the results of three very fruitful techniques in the study of nucleic acid structure, nuclear magnetic resonance, ultraviolet absorption and optical rotatory dispersion. Experimentally, they have much in common. They are all solution measurements. That means that temperature, pH or solvent composition can be varied continuously in order to observe changes in these conformation dependent physical properties caused by small changes in the environment.

We have already mentioned that evidence from nuclear magnetic resonance (NMR) measurements was instrumental in assigning a "stacked" structure to aggregated bases in solution. The proton resonances of the non-exchangeable protons of the purine bases are shifted upfield in concentrated aqueous solutions[98,128,217,219]. This results from the interaction of a proton with the $\pi$ electron system of the adjacent base. If the bases were vertically "stacked" one would expect an upfield shift[89]. If the bases were aggregating in a coplanar manner, then a down-

field shift would be expected. Indeed, this downfield shift is seen in purine and pyrimidine derivatives in solvents such as chloroform and dimethylsulfoxide[94] where infrared absorption studies show coplanar hydrogen-bonded interactions[70,100,101].

Quantitative comparisons among all bases of differing tendencies to "stack" are not possible because the degree of aromaticity varies considerably for purine and pyrimidine bases. The purines are quite aromatic while the pyrimidines, especially uracil and thymine are much less so. Proton resonances have the smallest concentration dependence in the case of uracil. It is difficult to say whether this is caused by very little aggregation or by the small interaction between protons on one base and the $\pi$ electrons on another.

Measurements of the optical properties of nucleic acids have given much insight into the interactions between adjacent nucleotides. Purine and pyrimidine bases have strong absorptions in the ultraviolet region near 260 mμ[205]. The magnitude of this absorption depends on the local environment of the base. The extinction coefficient (at 260 mμ) of a nucleotide in native DNA is reduced by almost a factor of two compared with the extinction coefficient of that nucleotide free in solution[7,105,159,188]. This decrease in extinction coefficient at a given wavelength is called hypochromicity while the decrease in oscillator strength (the extinction coefficient integrated over an entire absorption band with respect to

frequency) is called hypochromism[119]. Experimental studies of single-strand oligonucleotides (with no base-base hydrogen bonding)[126] as well as theoretical investigations[46,160,189] have shown that in general, chromophores, close enough to one another so that there is appreciable coulombic interaction between them, will exhibit a change in the intensity of absorption, either hyperchromic or hypochromic. Hypochromism of single-strand nucleic acid material in aqueous solution indicates that their bases are near one another under these conditions. However when the temperature of the solution is raised, or when denaturants such as alcohols or urea are added, this hypochromism decreases, apparently going to zero in the limit of high temperature or high concentration of denaturant in many cases[3,16,43,106,147,203,206]. This is equivalent to saying that the ultraviolet absorptions of oligo- and polynucleotides go toward those of free nucleotides under denaturing conditions. In the case of dinucleoside phosphates, it implies that raising the temperature or adding a denaturant increases the distance between bases, breaking up the "stacked" structure in favor of one in which the distance between bases is limited only by the covalent bonds between them.

Optical rotatory dispersion (ORD) and circular dichroism (CD) of single-strand oligonucleotides and polynucleotides, especially dinucleoside phosphates, give results which are quite similar. (Since ORD and CD can be converted from one to the other by the Kronig-

Kramers transforms[132], two linear integral transforms, one can speak of these qualitative results almost interchangeably.) The ORD and CD of dinucleoside phosphates are very different from those of their component nucleosides and nucleotides[14-16,26,209-211]. This is a result largely of the interaction between bases[26,27,209]. Again, when the temperature is raised or when denaturing solvents are added the ORD or CD of an oligonucleotide or polynucleotide tends to go to that of its component nucleotides, indicating a large decrease in the interaction between bases[14-16,43,79,165,197].

Base composition and sequence can cause large variations in the ORD or CD of dinucleoside phosphates. The ORD of adenylyl (3'-5') adenosine (ApA) is about an order of magnitude larger than that of adenosine (A) and quite different in shape while the ORD of UpU is quite similar to that of U with only a fractional increase in magnitude[210]. This is further evidence which makes it necessary to include base composition and sequence dependence in any acceptable theory of the structure of dinucleoside phosphates in solution.

In spite of the fact that there seem to be so many different experiments, there is really very little information about the structures of dinucleoside phosphates with which our conclusions must agree. The increase in base-base interaction with decreasing temperature or decreasing concentration of denaturants indicates that the low temperature limit (i.e. ordered form) of dinucleoside phosphate structure must be "stacked". More specifically,

the bases must be close to one another, probably parallel with no solvent between them, in a conformation somewhat like that of two adjacent nucleotides of the same strand of a helical DNA. It is very difficult to be more specific. The limit of the conformation at high temperatures or in denaturing solvents such as alcohols or urea solutions is one in which the bases are farther away from one another, so that any coulombic interactions between them (in the ground or excited states) have diminished greatly. Again, it is impossible to be more specific with only this evidence at our disposal. One attractive suggestion is that the bases move freely through the solution, constrained only by the covalent bonds connecting them. However attractive this may seem, there is no direct evidence supporting it.

In addition, models must agree with the energetic parameters for the "stacking" process as determined by both calorimetric[66,179] and statistical mechanical[40] methods. Because the systems are so different, a great many assumptions must be made in order to compare energies obtained from the helix-coil transition or aggregation of bases and base analogues in solution with any value of the energy of "stacking" for dinucleoside phosphates calculated using a model we might propose. It is unfortunate that critical comparisons will be impossible because of these difficulties.

Any realistic model must also allow for differences in the properties of dinucleoside phosphates differing in

base composition and sequence. It would be useful if we were able to correlate this sequence dependence with a single property or small number of properties of each base.

PHYSICAL ORIGINS OF THE PROPERTIES TO BE MEASURED

The three physical measurements which seem most useful in studying the conformations of dinucleoside phosphates were ORD, ultraviolet absorption and NMR. They enable us to investigate dinucleoside phosphates in solution where we can apply small perturbations such as changes in temperature, pH or solvent composition. In addition all three methods are good probes for the "unstacking" process in dinucleoside phosphates. The ORD of dinucleoside phosphates changes by as much as 10 fold with varying temperature (Figure 1). The ultraviolet absorption changes by as much as 13% over a similar range (Figure 2). The positions of the proton magnetic resonances of the non-exchangeable base protons can change by more than 0.3 ppm depending on the experimental conditions[21]. The experimental precision obtainable with commercially available instruments is roughly one or two orders of magnitude better than the largest of these effects, making the measurements tractable and attractive.

The ORD, ultraviolet absorption and NMR all have different geometric dependences. They give slightly different views of the changes in the conformations of

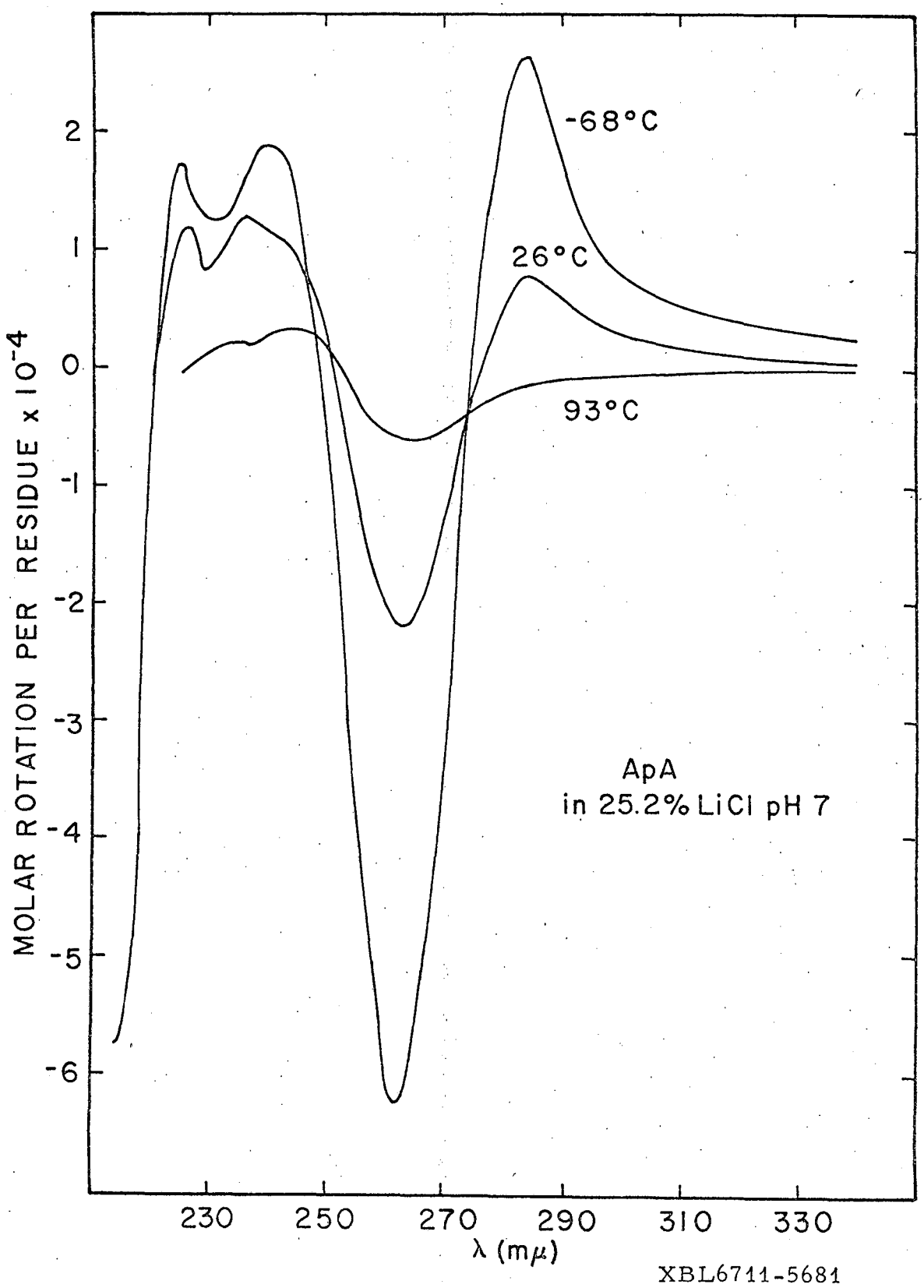



Figure 1 ORD of ApA in 25.2% LiCl solution, -68°, 25°, 93°C.

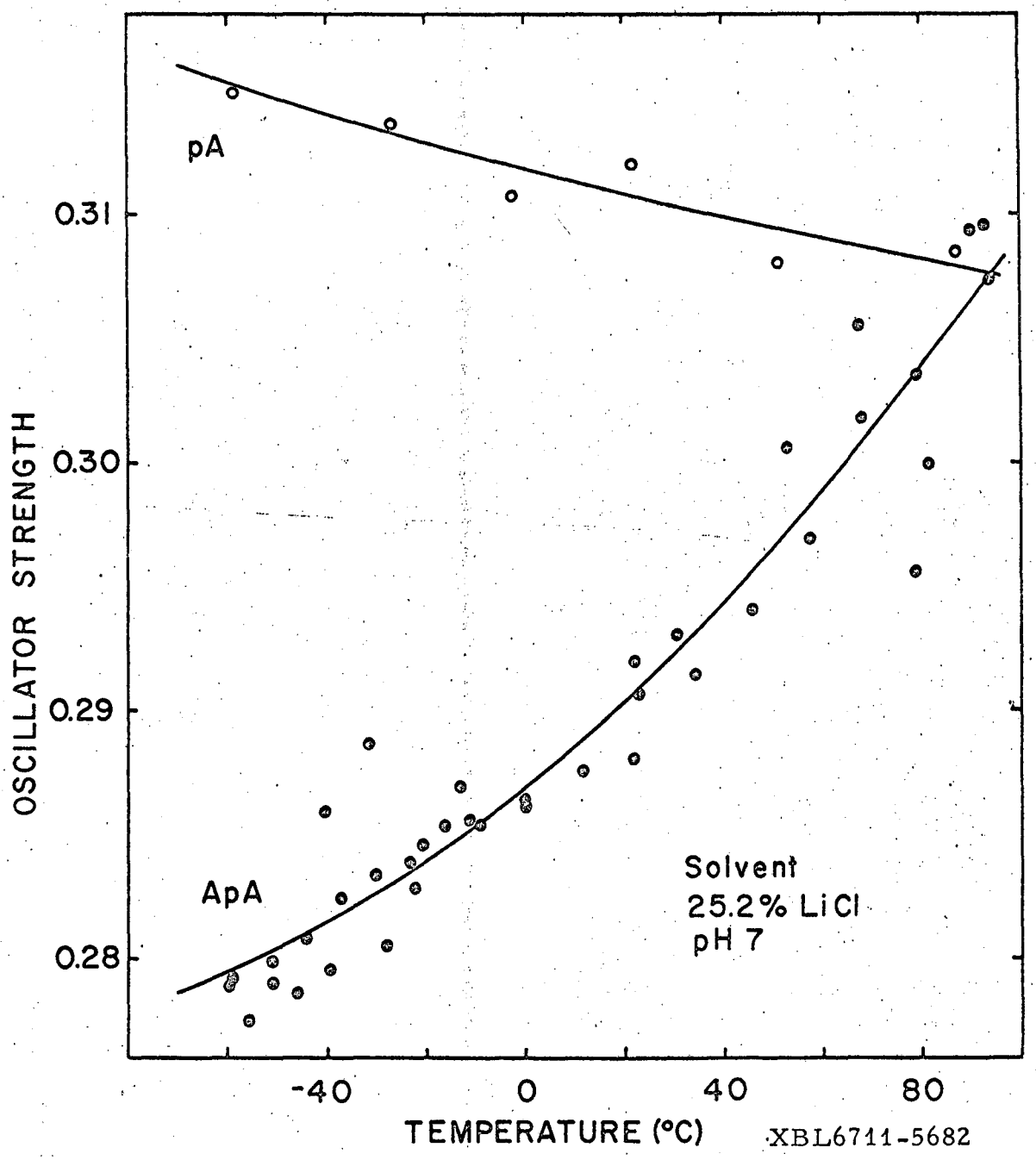



Figure 2 Variation of the oscillator strength of ApA in 25.2% LiCl solution as a function of temperature.

dinucleoside phosphates, giving additional insight into the problem. In order to extract the maximum amount of information from these measurements it is very important that we understand the physical bases of the effects.

OPTICAL ROTATORY DISPERSION

The region of interest in the ORD of dinucleoside phosphates is that near the base absorptions which occur in the near ultraviolet (near 260 mμ). A large change of the ORD in this region is caused by interactions between the two bases, in particular the interaction of the 260 mμ transition of one base with the electronic transitions of the other base[210,211]. This includes the interaction of the 260 mμ transition of one base with the 260 mμ transition of the other. In the case of ApA this is the dominant contribution[209,210]. This type of ORD, where the magnitude of the positive and negative Cotton effects in the region of an absorption band are equal, has been called "conservative"[26]. In this case (with identical bases) the magnitude of the rotation will be proportional to $V(\vec{r}_{12}\cdot\vec{\mu}_1 x\vec{\mu}_2)$ where $\vec{r}_{12}$ is the vector connecting the point transition dipole of one base with the point transition dipole of the other base, $\vec{\mu}_1$ and $\vec{\mu}_2$ are the point transition dipoles for the 260 mμ (ground to first excited) transitions of the first and second bases respectively and V is the matrix element $\langle\psi_{10}\psi_{21}|\mathcal{H}|\psi_{11}\psi_{20}\rangle$ where $\mathcal{H}$ is the Hamiltonian (only coulombic interactions between bases need be considered because the bases are

far enough away from one another so that interactions caused by the overlap of the wave functions of the two bases can be neglected), and $\psi_{ij}$ is the wave function of the $i^{th}$ base in the $j^{th}$ excited state. If we assume that the coulomb interactions in V can be approximated by a dipole-dipole interaction then V will have an $r^{-3}$ dependence on the distance between bases. The important point of this treatment which should be remembered is that the dependence of the magnitude of the rotation of a dinucleoside phosphate like ApA is approximately proportional to $r^{-2}$ and has a strong angular dependence.

In other cases (e.g. CpC) the positive and negative Cotton effects near 260 mμ are not equal in magnitude[209,211]. This "nonconservative" ORD is probably caused by interactions of the transitions near 260 mμ on one base with the transitions of the other base which are higher in energy. There may also be interactions involving magnetic transition dipoles and permanent electric dipole moments[120]. Unfortunately, experimental data on the electronic transitions of purine and pyrimidine bases (especially the transition orientations) for the transitions near 260 mμ are quite limited. The situation deteriorates rapidly farther into the ultraviolet. This, combined with the fact that the best approximate wave functions for molecules as complicated as purines or pyrimidines are not particularly reliable, makes it difficult to calculate the ORD of dinucleoside phosphates. In addition, accurate treatment of solvent effects is very difficult[216] and has not been attempted

for molecules like dinucleoside phosphates yet. Calculations of the "conservative" term of the ORD of dinucleoside phosphates are close to the experimentally measured spectra in a few cases (e.g. ApA)[27,220]. However, the magnitude of the calculated ORD is generally not too reliable. In some cases even the qualitative shape does not agree with experimental results. Hopefully, the future will see accurate calculations of the ORD of dinucleoside phosphates as a function of conformation, enabling investigators to rigorously test models of the "stacking" process. Even without this information, the knowledge we do have about the origins of the ORD will aid us greatly in interpreting experimental results.

ULTRAVIOLET ABSORPTION

Hypochromism in dinucleoside phosphates results from the interaction of a transition on one base with other electronic transitions on the other base[189], with no contribution from the interaction of that transition on one base with the analogous transition on the other base[46]. However there can be appreciable changes in the band shape caused by such an interaction. In the case of a dimer consisting of two identical chromophores, there will be two exciton transitions for each one in the monomer[191]. There can be a "borrowing" of intensity from one band by the other. In the case of dinucleoside phosphates the exciton splittings (energy difference between two exciton transitions) are considerably smaller than the electronic absorption band widths[25]. The exciton transitions will

not be resolved in the absorption spectrum and integration over what appears to be a single absorption would actually include both exciton absorptions. Any exchange of intensity between these transitions will cancel. However, if one measures the extinction coefficient at just one wavelength (hypochromicity) he may observe changes caused by exciton interactions, or other changes in the band shape. (We will see later that the band shape of ApA does change appreciably as the temperature is varied.) Just this consideration shows that different optical properties depend on different interactions. The ORD of ApA is dominated by the interaction of the 260 mμ transition on one base with the 260 mμ transition on the other base. This interaction can have no effect on the hypochromism. However, because there may be changes in the shape of the 260 mμ absorption band there may be differences between the hypochromism (decrease in oscillator strength) and hypochromicity (decrease in extinction coefficient at one wavelength). Care should be taken to avoid confusing them.

The major contribution to the hypochromism of dinucleoside phosphates near 260 mμ originates in the interaction of the 260 mμ transition on one base with the transitions of the other base farther into the ultraviolet[49,189]. If a multipole expansion is used for the interaction between transitions the first term will be a point dipole-point dipole interaction which has a base-base distance dependence proportional to $r^{-3}$. There are a large number of transitions

in the ultraviolet with which the 260 mμ transition can interact. Since the orientations of these transitions are varied the angular dependence of the hypochromism is not as strong as that of the ORD. As the angle between the 260 mμ transition and a transition on the other base (farther into the ultraviolet) changes so that the interaction between them causes a smaller hypochromic effect, the orientation of the 260 mμ transition and another transition in the far ultraviolet may change so that its hypochromic contribution increases. This cancellation of changes in the terms contributing to hypochromism make the hypochromism less sensitive to the angular orientation of the two bases of a dinucleoside phosphate than the ORD. . . . If transitions are above one another (as in "stacked" bases) the long wavelength transition (i.e. the one near 260 mμ) will never be hyperchromic[133]. Therefore we expect to find the absorption band of a dinucleoside phosphate near 260 mμ to be hypochromic with an $r^{-3}$ base-base distance dependence and a weaker angular dependence than the ORD.

A classical approach to the problem of the origin of hypochromism treats it as an interaction of a chromophore with the local electric field caused by a polarizable body. The electric field of the radiation induces a dipole in the polarizable body which then has its own electric field. If this field partially cancels the electric field of the light at the site of the chromophore then hypochromism

results, the smaller electric field of the light at the chromophore resulting in less absorption. In the case of dinucleoside phosphates, the 260 mμ transition dipole interacts with the polarizability of the other base (caused by the transitions farther in the ultraviolet on that base). The variation of the polarizability of a purine or pyrimidine probably does not change by more than a factor of two in any direction in the plane of the base. The angular variation of the hypochromism (assuming the distance between bases remains constant and the bases remain parallel) should not be larger than this factor of two. This is analogous to the argument used earlier which said that since the ultraviolet transitions were distributed in different directions hypochromism should have a small angular dependence.

Qualitatively, the hypochromism should have a smaller angular dependence than the ORD (which can even change sign) and should vary approximately as $r^{-3}$. (The ORD varies approximately as $r^{-2}$.) Complete analysis of experimental data requires the calculation of the physical property measured as a function of molecular conformation. Unfortunately, quantitative calculations of hypochromism are very difficult[49, 160,189]. One major obstacle is the problem of knowing details of the transitions in the ultraviolet. For some cases the ORD is dominated by interactions of the transitions near 260 mμ with one another. We have seen that this can never be the case for hypochromism. That means that accurate calculations of the hypochromism of dinucleoside phosphates

requires a knowledge of the transitions not only at 260 mμ, where our knowledge is meager, but also knowledge of the transitions further in the ultraviolet, where the situation is far worse. In addition, there are electronic transitions in the solvent which may interact strongly with the purine or pyrimidine transitions. This effect is particularly difficult to calculate for the reasons previously stated as well as the fact that very little is known about the solvent structure around molecules like dinucleoside phosphates. Since the bases of a dinucleoside phosphate are exposed to the solvent (much more so than in DNA) this effect may be large.

NUCLEAR MAGNETIC RESONANCE

NMR experiments measure something quite different. In these experiments one does not observe the interaction between excited electronic states. The experiments of interest measure the positions of the resonances of non-exchangeable protons on the base (e.g. $H_2$ and $H_8$ of adenine). The positions of these resonances depend critically on the position of the other base of the dinucleoside phosphate[21,33,35,169]. When an aromatic ring is placed in a magnetic field the $\pi$ electron movement becomes polarized. In very simple terms, a "ring current" is established. This "ring current" generates its own magnetic field which either adds to or subtracts from the external magnetic field[89]. If a proton is located directly above or below the aromatic ring then the magnetic field generated by the "ring current" opposes the external magnetic field. A higher external

magnetic field is needed for resonance; the resonance is shifted upfield. On the other hand, if the proton is in the same plane as the aromatic ring (as is the case with most protons covalently bound to aromatic rings) the magnetic fields enhance one another. A lower external magnetic field is needed for resonance; the signal is shifted downfield.

The effect one observes in dinucleoside phosphates, nucleosides or bases in solution is an upfield shift of the resonances of the non-exchangeable hydrogen on the aromatic rings with increased "stacking" of bases. In a dinucleoside phosphate with two similar bases (e.g. ApA) this gives an unusually sensitive measurement of conformation. The two bases are geometrically non-equivalent (caused by the differences between the 3' phosphate ester and the 5' phosphate ester) and therefore magnetically non-equivalent. Each proton of one base interacts with a different part of the other base. This is greatly enhanced in "stacked" conformations. The differences between the two proton resonances can be measured very accurately. This difference gives a very sensitive method of measuring the differences of environment seen by the two bases.

Unfortunately, the aromaticity of the nucleic acid bases varies greatly. Although the purines display large "ring currents" the pyrimidines, especially uracil and thymine have very small "ring currents"[169]. In the case of pyrimidine bases it is difficult to differentiate between the case where the bases are "stacked", but no "ring current" effect is seen because the "ring current" is too small and

the case where the bases simply do not "stack".

The interaction of the aromatic electrons with a proton in a magnetic field can be treated on a very simple level as two point dipoles interacting. The electron "ring current" (which is equivalent to a magnetic dipole) interacts with the proton's spin (which is a magnetic dipole). Therefore the dependence of the interaction on the distance between bases should be proportional to $r^{-3}$. Of course, when r is of the same size as the dimensions of the aromatic ring, the "ring current" can not be considered to generate a simple magnetic point dipole. More rigorous calculations of this effect have been made, considering both the geometry of the aromatic ring and the basic quantum nature of the problem[89]. Treatments of molecules as complicated as dinucleoside phosphates are just starting to be made[21].

The measurement of ORD, ultraviolet absorption and NMR of dinucleoside phosphates offer different ways of investigating the conformation of dinucleoside phosphates in solution. They are all sensitive to interactions between the two bases but measure slightly different quantities. Using these techniques we will be able to test some models and develop ideas about the conformation of dinucleoside phosphates in solution, as well as the forces responsible for these structures.

MODELS

With the exception of certain types of X-ray diffraction studies, analysis of physical measurements requires a

choice of a model in order to determine a structure. If the data are not consistent with the model then the model must be discarded or revised. The more data agree with a model, the more credible it becomes.

Before discussing the physical measurements which can yield useful information about the conformation of dinucleoside phosphates in solution it would be advantageous to mention some models which are compatible with the qualitative results already reviewed. They will serve as guides in understanding the conformations of dinucleoside phosphates under various conditions as well as help suggest the measurements which might enable us to distinguish between models.

TWO-STATE MODEL

The simplest model used to analyze the properties of dinucleoside phosphates and other oligonucleotides (indeed the only model used up to this time) is the two state model[3,14,15,106,147,203,206]. This model assumes that a dinucleoside phosphate can have only one of two possible conformations, either the ordered "stacked" conformation or the disordered "unstacked" conformation. No conformations between these two exist. Variation of temperature or solvent composition changes the equilibrium between the two states. Lowering the temperature favors the ordered form. Raising the temperature or introducing denaturing solvents shifts the equilibrium toward the disordered form. The physical properties of the ordered form can be determined

by extrapolating these properties of the equilibrium mixture to the low temperature limit. Comparisons with properties of known structures or properties calculated from theory may help to determine the geometry of the "stacked" dinucleoside phosphate. Extrapolation to the high temperature limit will give the properties of the "unstacked" dinucleoside phosphate. Similar comparisons with known compounds as well as calculations based on theory may lead to the determination of the geometry of the disordered form.

$\Delta H^o$ and $\Delta S^o$ of the "unstacking" process can be determined by measuring the equilibrium constant for the "stacked" "unstacked" equilibrium as a function of temperature, then applying the Van't Hoff equation:

$$\frac{\partial \ln K}{\partial \frac{1}{T}} = - \frac{\Delta H^o}{R}$$

together with:

$$- RT\ln K = \Delta H^o - T\Delta S^o$$

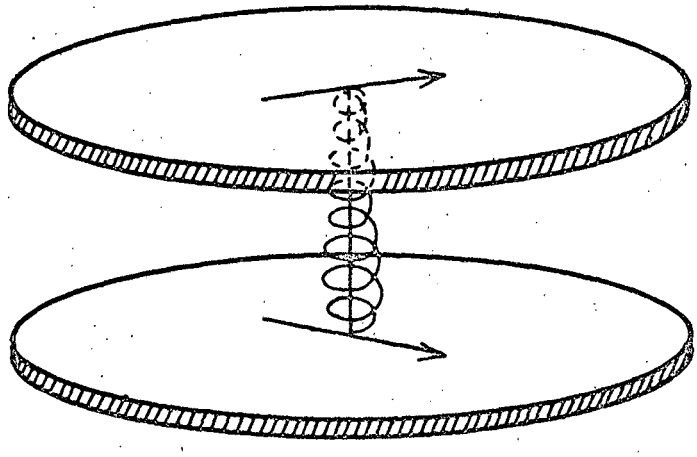
MULTI-STATE MODEL

A logical extension of the two-state model is the multi-state model. In this model a finite set of possible conformations exists, ranging from "stacked" to "unstacked", each with its own physical properties and its own $H^o$ and $S^o$. Analysis of the data, using this model, is almost identical to the analysis using the two-state model in

principle. In practice it is much more difficult. In the case of the dinucleoside phosphates there is at this time no method of determining the population distribution for the intermediate states. This is necessary in order to treat the multiple-equilibrium problem which is necessary for the determination of the thermodynamic variables of each state. In spite of this fact it is useful to keep this model in mind as it will help provide some insight into the meaning (or lack of meaning) of $\Delta H^{o}$ and $\Delta S^{o}$ determined by using the two-state model, should the two-state model prove to be incorrect.

OSCILLATING DIMER MODEL

Up to this point we have mentioned two models having discrete states. At the other extreme is the situation where a dinucleoside phosphate can exist in a continuum of states. Instead of discrete energies there would be an energy surface (in a multi-dimensional space). In one very simple model (Figure 3) the bases of a dinucleoside phosphate are considered to be two circular discs directly above one another[67,116]. A harmonic torsional spring connecting the centers of the two discs defines the potential energy. The discs can oscillate but can not move away from one another. In the limit of low temperature there would be no oscillations (in the quantum mechanical case, just zero point oscillations) from the equilibrium position and the dinucleoside phosphate would be "stacked". Increasing the temperature




$$\Phi = \Phi_o e^{-(\hbar/\omega m)\coth(\hbar\omega/2kT)}$$

at high temperatures

$$\Phi = \Phi_o e^{-2RT/\kappa}$$

hypochromism is constant

Figure 3 Representation of the oscillating dimer model.

XBL6711-5683

would increase the magnitude of the oscillations. In the limit of very large oscillations the dinucleoside phosphate would become "unstacked". Denaturing solvents would have the effect of reducing the force constant of the spring, allowing larger oscillations at the same temperature (i.e. at the same kinetic energy). Quantitative analysis of the data using this model would give a force constant ($\kappa$) for the torsional spring and the moment of inertia of the effective oscillator, rather than $\Delta H^o$ and $\Delta S^o$ for the disordering process. The energy required for a displacement of $\theta$ radians from the potential energy minimum is $\kappa\theta^2$. If a displacement of $\theta_u$ is required for "unstacking" a dinucleoside phosphate then the energy of "unstacking" would be $\kappa\theta_u^2$.

Using this model, the temperature dependences of some properties of dinucleoside phosphates can be calculated[67,116]. If the resulting ORD from the interaction of the two bases is considered to be that arising from the interaction of the point transition dipole (near 260 mμ) at the center of one base, with the analogous point transition dipole at the center of the other base, the contribution to the optical rotation caused by the interaction of the two bases

$$[\Phi]_{XpY} - \tfrac{1}{2}([\Phi]_X + [\Phi]_Y)$$

is

$$[\Phi] = [\Phi]_o \, e^{-(\hbar/\omega)\coth(\hbar\omega/2kT)}$$

in the quantum mechanical case. $[\Phi]$ is the rotation caused by interaction of the two bases; $\omega$ is the frequency of the oscillator; m is the moment of inertia of the disc; $\hbar$ is Planck's constant divided by $2\pi$; k is Boltzmann's constant; T is the absolute temperature; and $[\Phi]_o$ is a constant, proportional to the low temperature limit of the optical rotation. The classical (high temperature) limit of this expression is:

$$[\Phi] = [\Phi]_o \, e^{-2kT/\kappa}$$

where $\kappa$ is the force constant of the spring.

One feature of this model is that it predicts different temperature dependences for different properties. If the hypochromism is considered to arise from the 260 m$\mu$ transition interacting with the polarizability of the other base, and the polarizability is constant in the plane of the base then there will be no change in the hypochromism with changing orientation (or changing temperature). The variation of the polarizability in the plane of a base is probably less than a factor of 2 making the hypochromism slightly temperature dependent and leaving an appreciable hypochromism in the high temperature limit. (This is found in the case of ApA.)

However, the experimental hypochromism of ApA is found to be temperature dependent. The oscillating dimer model (which takes into account only those torsional oscillations of the bases which do not take them away from one another) can not account for this type of behavior. If other modes of motion are introduced then this type of model will predict a strongly temperature dependent hypochromism. In

our discussion of the origins of the ORD and hypochromism we mentioned that the ORD has a strong angular dependence and a dependence on the distance between bases of approximately $r^{-2}$. On the other hand, the hypochromism has a very weak angular dependence but varies as $r^{-3}$. This means that movement of the bases away from one another (while retaining the same angle between the 260 mμ transition dipoles of the two bases) will affect the hypochromism more strongly than it will the ORD. It is possible that the variation of the ORD with temperature originates largely in torsional modes of motion while the variation of the hypochromism with temperature is a result of increased separation of the bases. Unfortunately, the mathematics involved in treating this type of motion is considerably more complicated than mathematics of the torsional modes of motion[67]. No quantitative theory has been developed yet for treating the dependence of optical properties when such types of motion occur.

We realize at the start that this model is much too simple. However, it does serve an important pedagogical purpose by illustrating the type of alternative to the two-state model which will have to be explored in order to understand the temperature dependent properties of dinucleoside phosphates.

GENERALIZED ENERGY SURFACES

The approach in which a generalized energy surface is used is not limited by these overly simple assumptions. A more useful model would have a realistic geometry (based on crystallographic results) along with a realistic multi-

dimensional potential energy surface. This geometry and surface would then be tested by calculating (numerically or analytically) the temperature (or solvent) dependence of physical properties and then comparing these predictions with experimental observations. The choice of realistic geometries of the bases, sugars and phosphate from existing X-ray crystallographic data is feasible. The choice of a potential is very much more difficult. The second part of this thesis is devoted to this problem.

Consideration of these four models of the temperature and solvent dependence of the properties of dinucleoside phosphates suggests experiments which must be done in order to test these models. Properties of dinucleoside phosphates must be measured under varying conditions (e.g. different temperatures and different solvents). The more properties measured and the wider the range of conditions used, the more reliable the check of the model will be. Therefore we have measured some temperature dependent properties of a number of dinucleoside phosphates.

## MATERIALS AND METHODS

### NUCLEIC ACID MATERIAL

A, C, Ap, pC, ApA, ApG, UpC, CpA, CpC, GpU, and GpC used in these experiments were Calbiochem A Grade. UpG, CpG and GpA were purchased from Gallard-Schlessinger Chemical Mfg. Corp. pA and pU were purchased from Pabst Laboratories. The ApA used in the low salt hypochromism

measurements, ApC, ApU, and some of the UpU was supplied by Dr. Myron Warshaw. Their preparation has been described elsewhere[210]. All of these compounds were used without further purification. Most of the dinucleoside phosphates we checked for impurities by high voltage paper electrophoresis and were found to be homogeneous[28]. UpA (Calbiochem A Grade) and UpU (Calbiochem B Grade) were found to contain appreciable impurities. They were purified on a 50 cm Dowex 1X2 ion exchange column. The UpA was eluted in a $10^{-3}$ to $10^{-2}$ M HCl gradient. It was lyophilized and stored as a dry powder. UpU was eluted in $10^{-2}$ M HCl with a 0-0.4 M $NH_4Cl$ gradient. The salt was removed by passing the solution through a Sephadex G-10 column. The UpU was then lyophilized and stored as a dry powder.

NON-BIOLOGICAL MATERIALS

Because the LiCl solutions used in these experiments were so concentrated we considered the possibility of a small transition metal ion impurity interacting with a dinucleoside phosphate and changing its properties. Di- and trivalent metal ions were removed by treating concentrated LiCl solutions with an excess of Chelex 100 chelating resin (Bio-Rad Laboratories) batchwise at least five times. Emission spectroscopy on dried samples of this LiCl indicated calcium between 10 and 20 ppm. Since the resin binds $Ca^{+2}$ more poorly than other di- and trivalent metal ions and the impurity levels of other metal ions were at least 20 times lower (with the exception of $Mg^{+2}$)

in the original LiCl sample, it is safe to assume that other metal ions will not be present in appreciable quantities. If the pK for the binding of $Ca^{+2}$ to dinucleoside phosphates is 1.8 as it is for the binding to adenylic acid[121], the fraction of dinucleoside phosphate bound to $Ca^{+2}$ would be about 1% which should not affect our experiments. Because so much $Li^{+}$ is present the binding is probably considerably weaker.

Tap distilled water was used for these experiments. Other compounds were reagent grade or the equivalent and used without further purification.

The dilute salt buffers used for ORD and ultraviolet absorption studies were 0.15 M $NaClO_4$, 0.02 M phosphate pH 6.75 or 0.15 M KCl, 0.01 M phosphate pH 6.7. No difference was observed in the properties of dinucleoside phosphates in these buffers. The concentrated LiCl buffers were 25.2% LiCl (g/g), 0.004 M tris-HCl pH 7.1.

For the NMR experiments it was necessary to eliminate as many protons from the solvent as was practical so that they would not swamp out the signals of the solute protons, present in low concentration. All solutes measured in water were dissolved in $D_2O$ of at least 99.7% deuterium (Bio-Rad) then lyophilized to remove exchangeable protons from the solute. This was done at least twice. The solute was then dissolved in $D_2O$ (ApA) or $D_2O$ with enough $Na_2CO_3$ dissolved to be sure the solution was not acidic (A, pA, and Ap). For measurements made at acid pH the material was

dissolved in $D_2O$ to which enough concentrated HCl had been added to make the solution 0.1 M. In order to eliminate the possibliity of depurination, spectra were taken within one hour after dissolving the ApA.

DETERMINATION OF CONCENTRATIONS

Concentrations of solutes for the NMR experiments were determined by weighing. Concentrations of ApA were 0.030 to 0.033 moles base/liter. The total concentration of nucleoside or nucleotides were 0.05 to 0.08 moles/liter.

Concentration of nucleosides, nucleotides and dinucleoside phosphates in the ORD and ultraviolet absorption studies were determined by measuring the ultraviolet absorption at room temperature. Extinction coefficients of the dinucleoside phosphates in dilute salt solutions were assumed to be the same as those reported in the literature[211]. Extinction coefficients in 25.2% LiCl solutions were determined by making a concentrated solution of the dinucleoside phosphate in buffered 25.2% LiCl. Aliquots were diluted with 0.004 M tris buffer pH 7.1 or buffered 25.2% LiCl to the same volume with a base concentration of approximately $5 \times 10^{-5}$ M. The solution diluted with tris buffer was sufficiently dilute in salt (less than 0.5 M LiCl in all cases) that the extinction coefficients for dilute salt concentrations could be used. Therefore the ratio of the extinction coefficients in dilute and concentrated LiCl solutions was the same as the ratio of optical densities. (see Table 1)

TABLE 1

| COMPOUND | LOW SALT* $\epsilon_{max}$ | LOW SALT* $\lambda_{max}$ | HIGH SALT $\epsilon_{max}$ | HIGH SALT $\lambda_{max}$ |
|---|---|---|---|---|
| pA | 15400 | 259 | 15500 | 261 |
| pU | 10100 | 262.5 | 9360 | 262.5 |
| C | 8900 | 273 | 9420 | 273 |
| ApA | 13400 | 258 | 13300 | 259 |
| UpU | 9800 | 261 | 7300+ | 262 |
| UpC | 8700 | 264 | 8740 | 266 |
| CpC | 8300 | 272 | 8920 | 271.5 |
| Estimated errors | | | ±50 | ±1.0 |

* Ref. 34

\+ ($\sim 10^{-4}$M base/liter)

INSTRUMENTS AND MEASUREMENTS

ORD measurements were made with a Cary 60 spectropolarimeter. The spectral bandwidth was less than 1 mμ in all cases. Samples were used for no more than four spectra. Baselines (the rotation of the buffer in the cuvette used for the sample) were run after the series of sample spectra. The cuvette was then rinsed and refilled with sample and more spectra were measured. Spectra were compared with at least two baselines measured at or near the temperature of the sample in order to check for possible spurious signals from the spectropolarimeter. Spectra were frequently measured to 600 mμ in order to check for base line shifts which frequently occur as a result of irreproducible positioning of the cell. Rotations are expressed as $[\Phi] = \frac{100\theta}{\ell c}$ where $\theta$

is the rotation in degrees, $\ell$ is the length of the cell in centimeters and c is the concentration in moles base per liter. No adjustments were made for the refractive index of the solvent. Corrections were made for the expansion of the solutions with increasing temperatures. Dilute salt solutions were assumed to have the density of pure water. The density of concentrated LiCl solutions is known as a function of temperature from 0°C to 100°C[68]. We obtained densities below 0°C by extrapolation; this led to a constant density of 1.155 g/cc below -20°C. In all cases these corrections amounted to no more than 3½%. Measurements of ORD were made using 1 or 2 centimeter cells. 2 centimeter cells were found to be more satisfactory because their longer bases gave them greater mechanical stability. Baseline shifts are greatly reduced by using 2 centimeter cells rather than 1 centimeter cells. The optical density of the sample at 260 mμ was between 0.5 and 1.1 for all work with dinucleoside phosphates. In some of the work with nucleosides and nucleotides the optical density was as high as 1.8.

Measurements on UpU (in dilute salt solution) and ApU were done in the 1 cm thermostatted cell supplied by Cary. The temperature was measured with a copper-constantan thermocouple placed in a glass capillary (filled with mercury for good thermal contact), which was in contact with the solution during the measurements. This system was found to be very

unsatisfactory. At extremes in temperature the cell has a tendency to leak. If the edges of the removable windows were greased to prevent this, the grease tended to streak across the windows at high temperatures, giving large anomalous rotations. In addition, the tubing carrying the thermostatting liquid had to be wired together. When connections came undone, flooding of the cell compartment and adjacent parts of the spectropolarimeter resulted. In order to avoid these difficulties this thermostatting system was quickly abandoned. Thermostatting for all other ORD measurements was done by placing a 1 or 2 cm cylindrical cell in an aluminum cell holder through which the thermostatting liquid was circulated. The details of this cell holder are described elsewhere[51]. For temperatures below 0°C it was helpful to cover this block with styrofoam insulation. Surgical rubber tubing was found satisfactory for carrying the thermostatting liquid to and from the cell holder. It has sufficient mechanical strength at high temperatures, without excessive stiffness at low temperatures. Tygon tubing and brass bellows were also tried but were found to spring leaks (especially at extremes in temperature) with surprising regularity. Periodic inspections and occasional replacement of the surgical rubber tubing were necessary in order to avoid mechanical failure and flooding of the spectropolarimeter. Connections inside the cell compartment between tubing and the cell holder and top of the cell compartment were made using Poly-flo connectors (Imperial Eastman). Outside the cell compartment Kwik-

connect connectors (Imperial Eastman) were used. They have the advantage of automatically shutting off liquid flow when a connection is broken.

The temperature of the sample was measured in one of two ways. Either the temperature of the thermostatting liquid was measured and converted to the sample temperature, using a previously determined calibration curve, or the temperature in a cell was measured directly with a copper-constantan thermocouple. This was done either immediately after a spectrum was recorded with the sample in the cell (after which the sample was discarded and the cell scrupulously cleaned in 50% $H_2SO_4$-50% $HNO_3$ in order to eliminate traces of transition metal ions), or immediately before the spectrum was taken (in which case an identical cell filled with water or alcohol into which the thermocouple was placed was mounted next to the sample cell in the aluminum cell holder). In all cases the temperature was known to be better than ±0.3°C and did not fluctuate beyond those limits. Temperatures below room temperature were maintained using refrigerated baths or by circulating methanol, cooled by Dry Ice or liquid nitrogen. Temperatures above room temperature were maintained using a Haake Model Fe[illegible] thermostatting bath. After the thermostatting bath had reached its regulating temperature it took the sample 5 to 15 minutes to reach thermal equilibrium (depending on the experimental set-up). At least an additional 5 or 10

minutes were allowed to pass before starting to measure spectra. In order to minimize evaporation above room temperature the cell was capped with rubber serum stoppers through which a small length of no. 27 hypodermic needle was placed. The serum stoppers were previously soaked in boiling 1 M KOH and rinsed in distilled water to leach out a material which absorbed in the ultraviolet. Hydrolysis of the dinucleoside phosphates at high temperatures was checked by measuring the ORD (or ultraviolet absorption) at room temperature both before and after having measured it at higher temperatures. In no case was the difference between the optical properties at room temperature measured before and after exposure to high temperature indicative of any hydrolysis.

Ultraviolet absorption measurements were made with a Cary 14 or Cary 15 spectrophotometer. Measurements not at room temperature were made in the thermostatted cell holder supplied by the manufacturers. Temperature control and measurement were done in the same way as they were for ORD measurements. The reference cell, containing the appropriate buffer was at room temperature. Therefore it was necessary to measure the difference in absorption of the buffer at room temperature and at the temperature of the sample.

The oscillator strength of an absorption band was determined by integrating the extinction coefficient using Simpson's rule, The optical densities of dilute salt solutions were read off the chart paper at 2.5 mμ intervals.

The integration was done using a computer program written by Dr. Myron Warshaw. Optical densities of other solutions were recorded at 0.5 mμ intervals using a Cary Digital System. The data were smoothed by fitting them, 25 at a time, to the best least squares cubic polynomial. The value of the smoothed optical density was taken to be the value of the polynomial at the $13^{th}$ wavelength of the 25 used for the fit. The smoothing and subsequent integration was done using a computer program written by Mr.S. Richard Jaskunas. For the 260 mμ bands of the bases the integration was done from long wavelengths to a cut-off wavelength at or near the minimum in the absorption spectrum. This cut-off wavelength varied from one dinucleoside phosphate to another, but was independent of temperature for any given compound. Since the near ultraviolet absorption bands of purines and pyrimidines are not completely resolved the choice of the cut-off wavelength is to some extent arbitrary. It is possible that errors may be introduced if the absorption bands change shape or shift with changing temperature.

NMR spectra were measured on a Varian Associates A-60 NMR Spectrometer. Measurements were made at 41°C unless otherwise noted. The temperature was determined by measuring the separation of the two proton signals of ethylene glycol, then determining the temperature using the calibration supplied by Varian. Temperatures other than 41°C were maintained using the V-6040 variable temperature controller made by Varian.

pH measurements were made at room temperature with either a Radiometer Type TTTlc pH meter or a Beckman Expandomatic pH meter. No corrections were made for high concentrations of $Li^+$.

## RESULTS AND DISCUSSION

### TWO-STATE MODEL

Because the change in the ORD is so much larger than that in the ultraviolet absorption or NMR most of our data will come from ORD measurements. We have measured the ORD of 14 dinucleoside phosphates in aqueous solutions over the range of approximately 0° to 85°C. (Figures 4-7) (The temperature dependence of the ORD of ApA was measured in this laboratory previously[210].) Using these data one can calculate an equilibrium constant as a function of temperature.

$$K(T) = \frac{\text{fraction "stacked"}}{\text{fraction "unstacked"}} = \frac{[\Phi](T)-[\Phi]_u(T)}{[\Phi]_s-[\Phi](T)}$$

All rotations are measured at one wavelength and $[\Phi]_s$ and $[\Phi]_u(T)$ are respectively the rotation of the "stacked" and "unstacked" forms of the dinucleoside phosphate and $[\Phi](T)$ is the rotation measured at temperature, T. Once the equilibrium constant is known as a function of temperature, thermodynamic parameters for the "stacked" to "unstacked" transition can be calculated.

In practice, one can make a reasonable estimate of the "unstacked" rotation. Since there is little base-base interaction the contributions to the optical rotation would

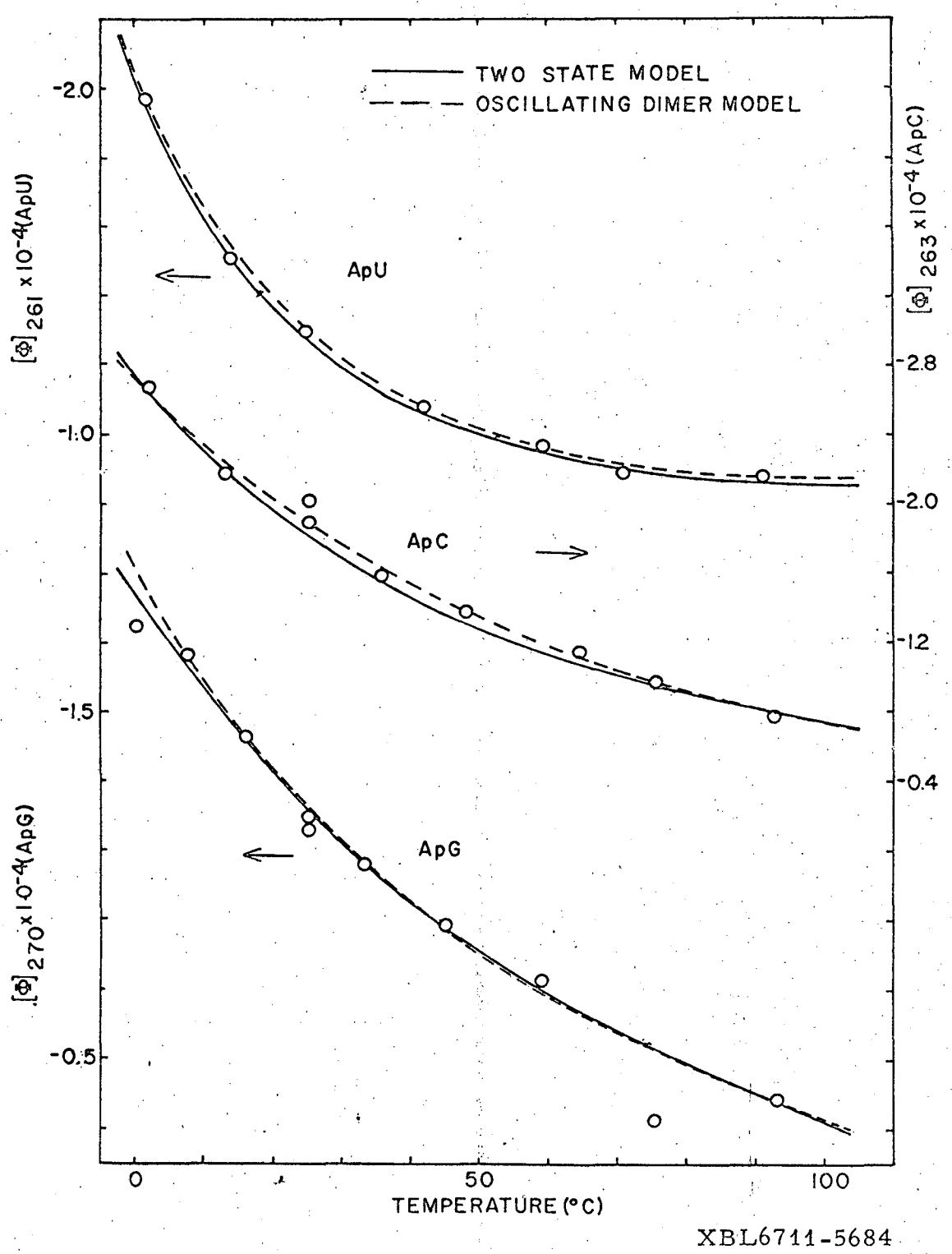



Figure 4 Temperature dependence of [Φ] for ApU, ApC and ApG.

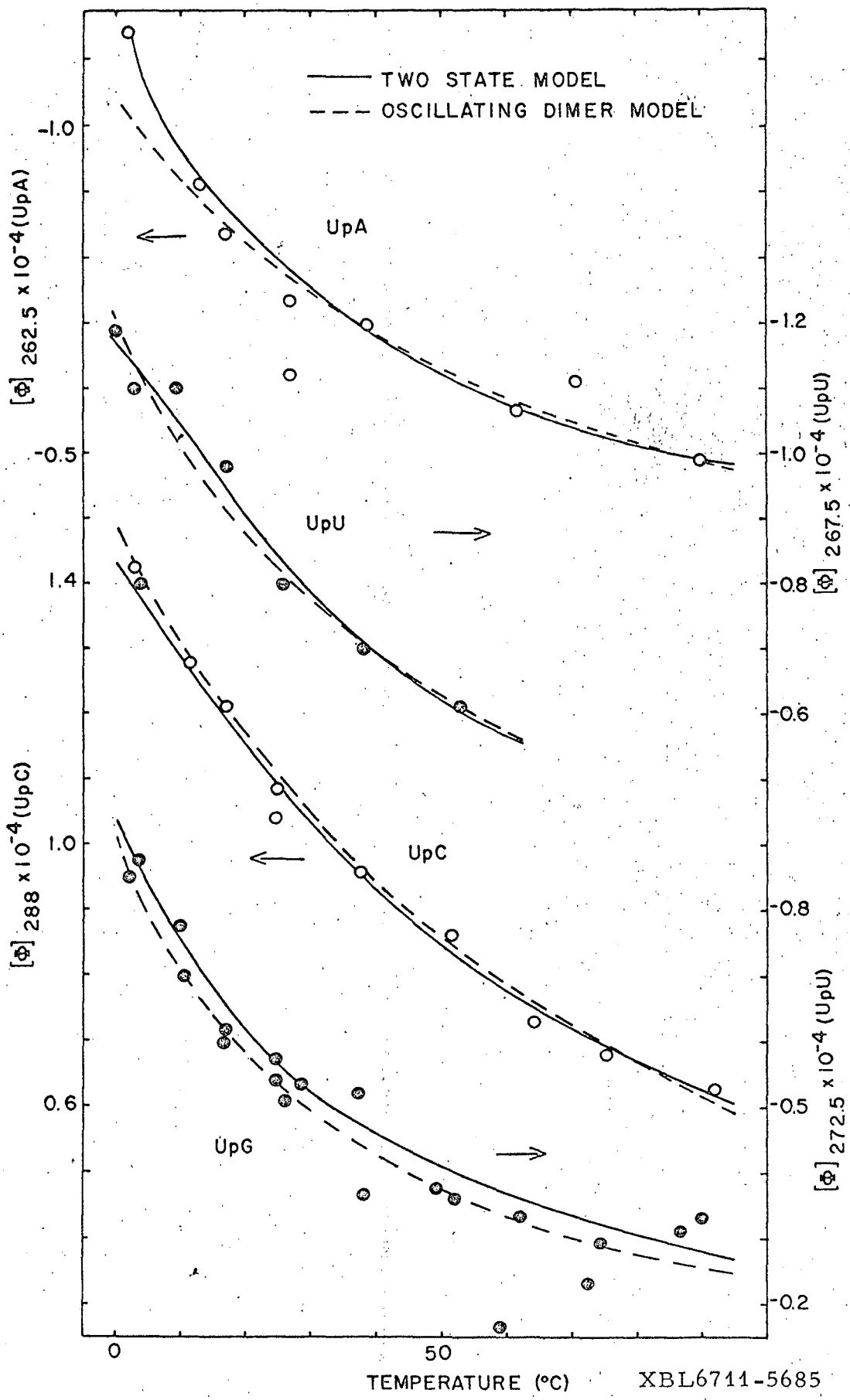



Figure 5 Temperature dependence of [Φ] for UpA, UpU, UpC and UpG.

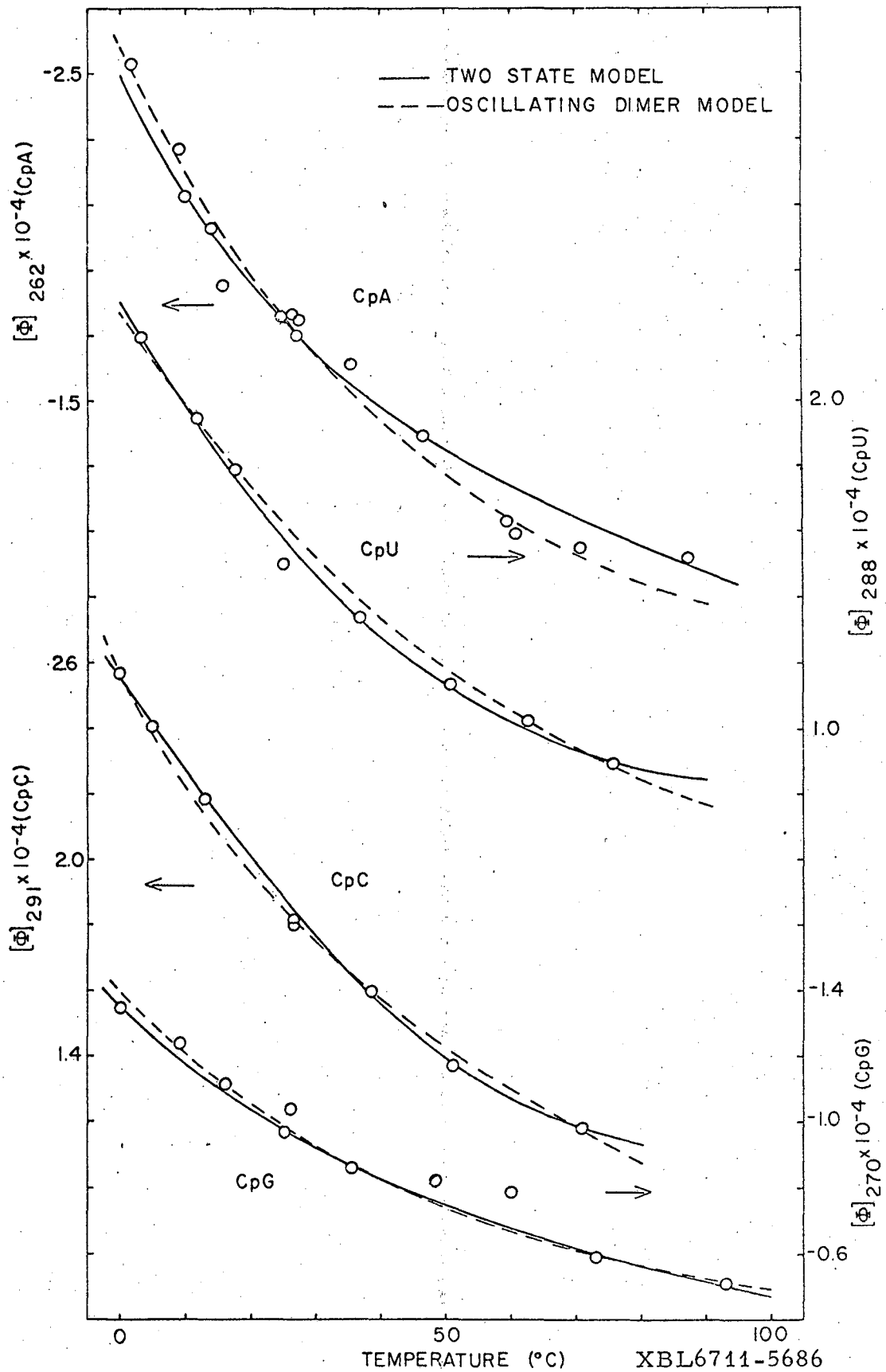



Figure 6. Temperature dependence of [Φ] for CpA, CpU, CpC, and CpG.

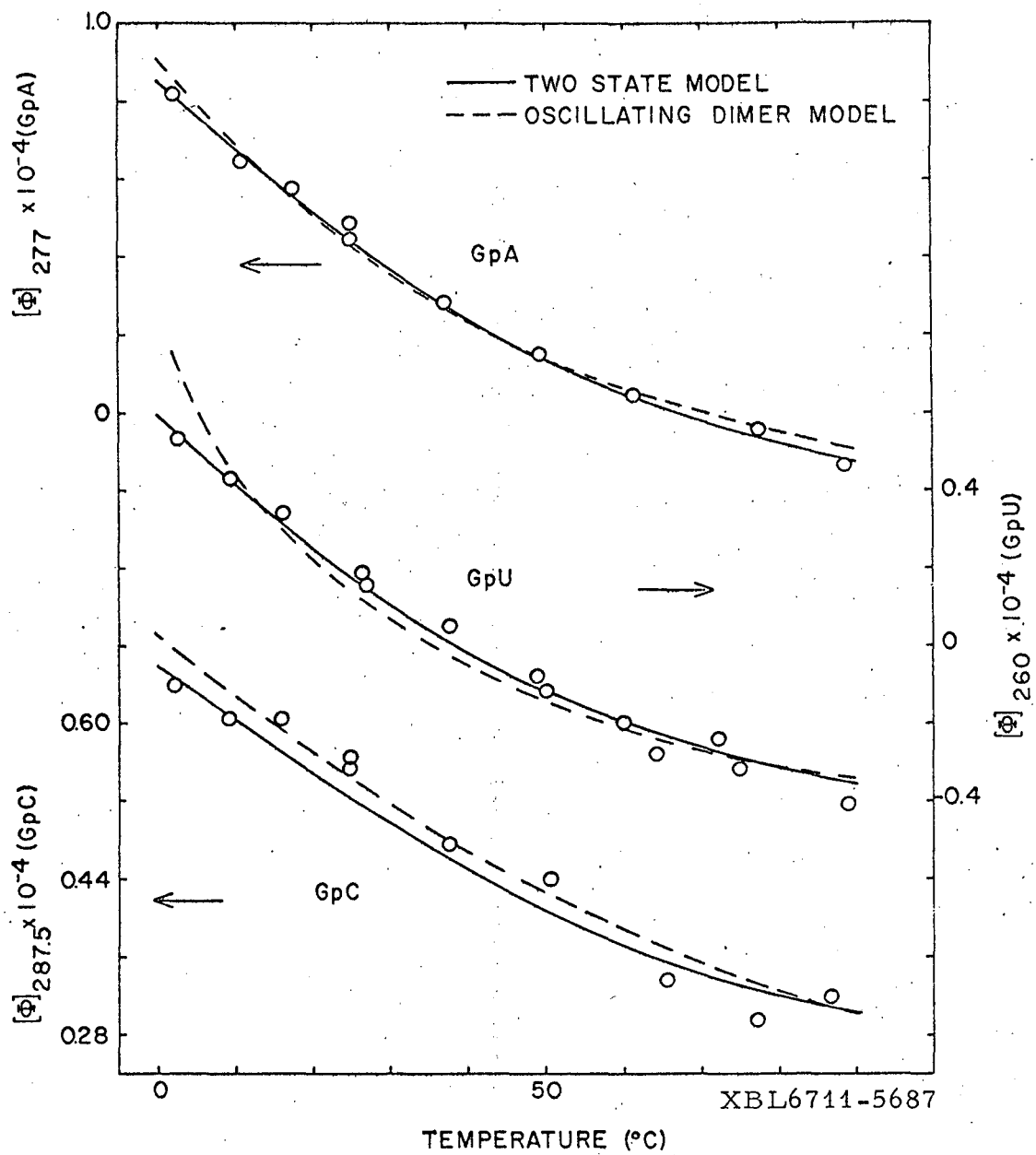



Figure 7 Temperature dependence of [ϕ] for GpA, GpU and GpC.

be those found in the free nucleosides or nucleotides. The extrapolated high temperature limit of the ORD of a dinucleoside phosphate does seem to be the sum of the rotations of its component nucleotides. The sum of the rotations of the component nucleotides was taken to be the rotation of the "unstacked" form.

However, nucleosides and nucleotides are not rigid molecules either. Since the bases are planar and therefore have no optical activity of their own, the optical rotation in the region of strong absorption by the bases (near 260 mμ) must result from an interaction of the 260 mμ transition with the asymmetric ribose. This interaction depends on the relative orientation of the base and sugar. Because the glycosidic bond (connecting the base and sugar) is a single bond, some rotation can take place around it. The population distribution of orientations will change as a function of temperature, with the conformations of higher energy becoming more populated at higher temperatures. The temperature dependence of the relative orientation of the base and sugar causes a temperature dependence of the ORD of nucleosides and nucleotides. Although the ORD of nucleosides and nucleotides remains qualitatively the same, the magnitude of the rotations decrease by as much as 25% going from 0° to 85°C. (Figures 8-13). These findings are in agreement with ORD measurements made in the visible region[195]. Therefore at each temperature we equate the rotation of the "unstacked" dinucleoside phosphate to the rotation of

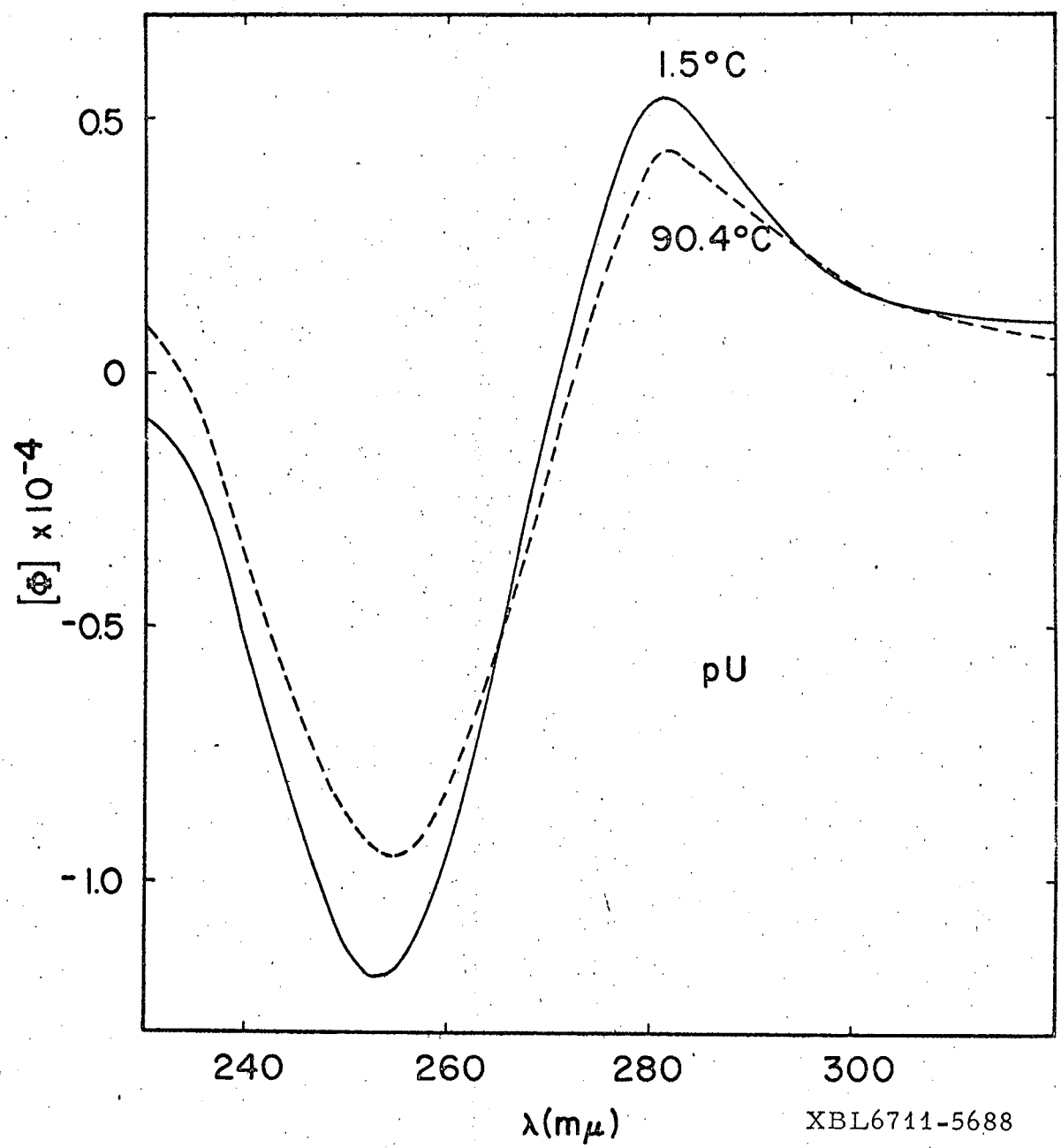



Figure 8 ORD of pU at 1.5° and 90.4°C.

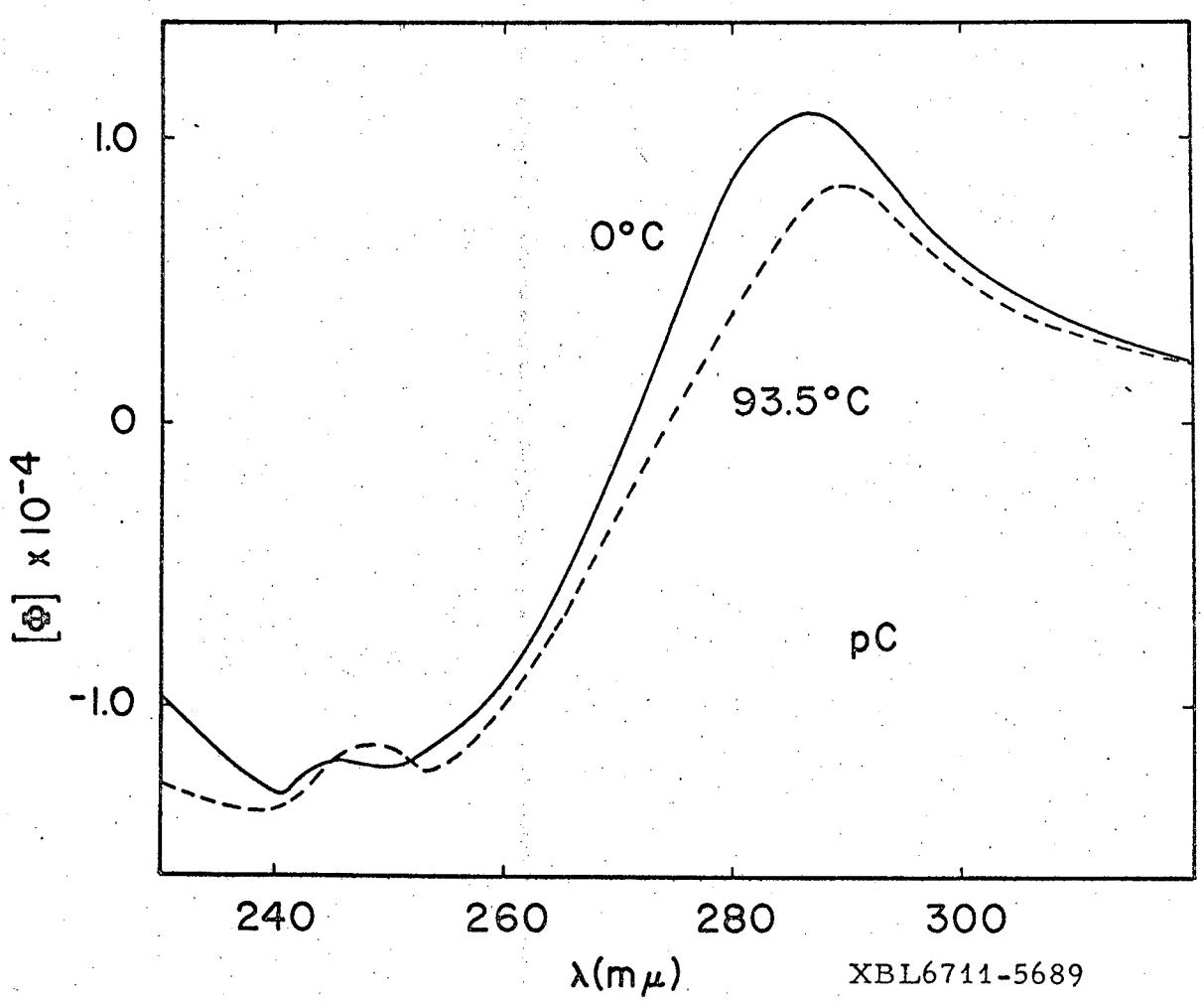
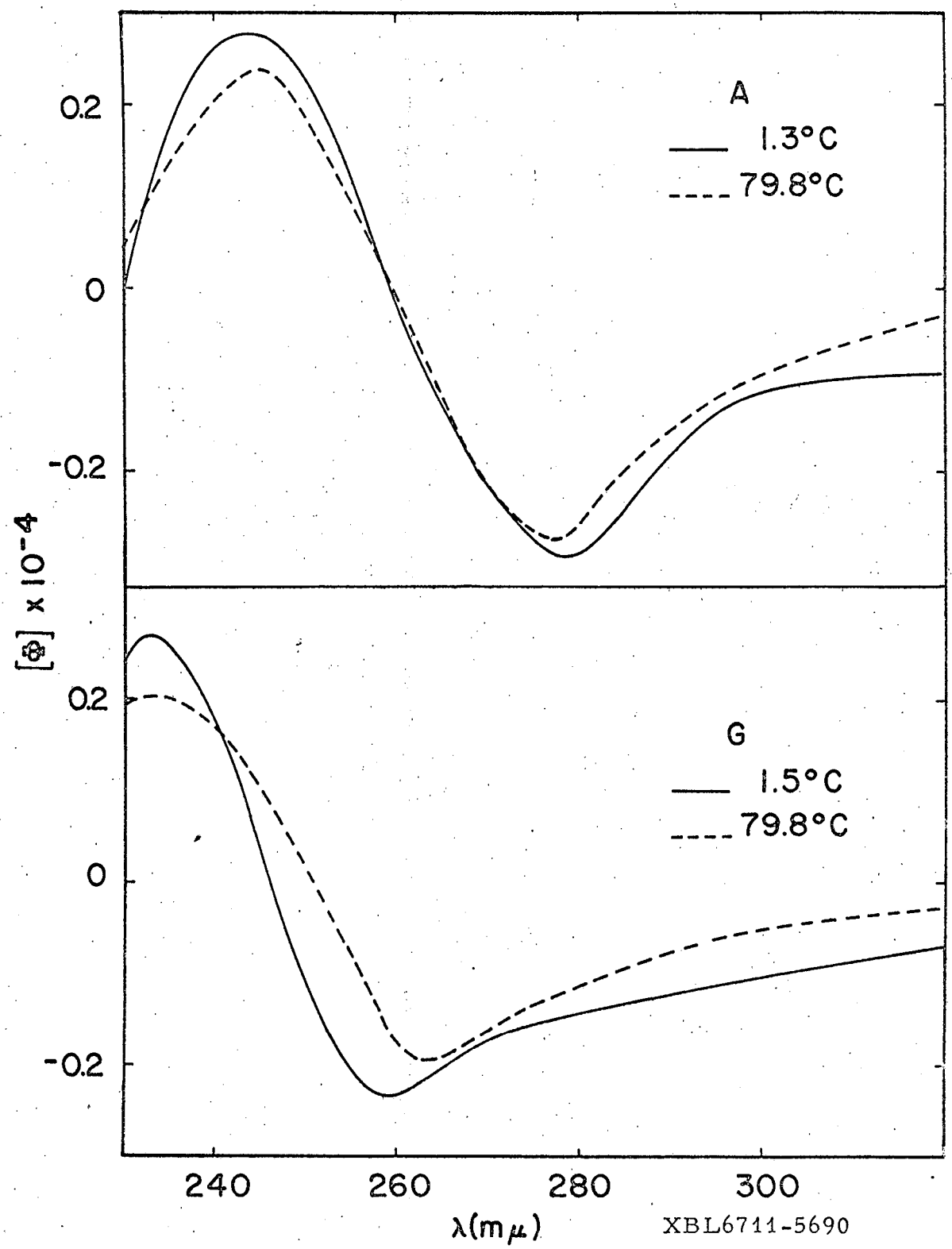



Figure 9 ORD of pC at 0° and 93.5°C.




Figures 10 and 11 ORD of A at 1.3° and 79.8°C.

ORD of G at 1.5° and 79.8°C.

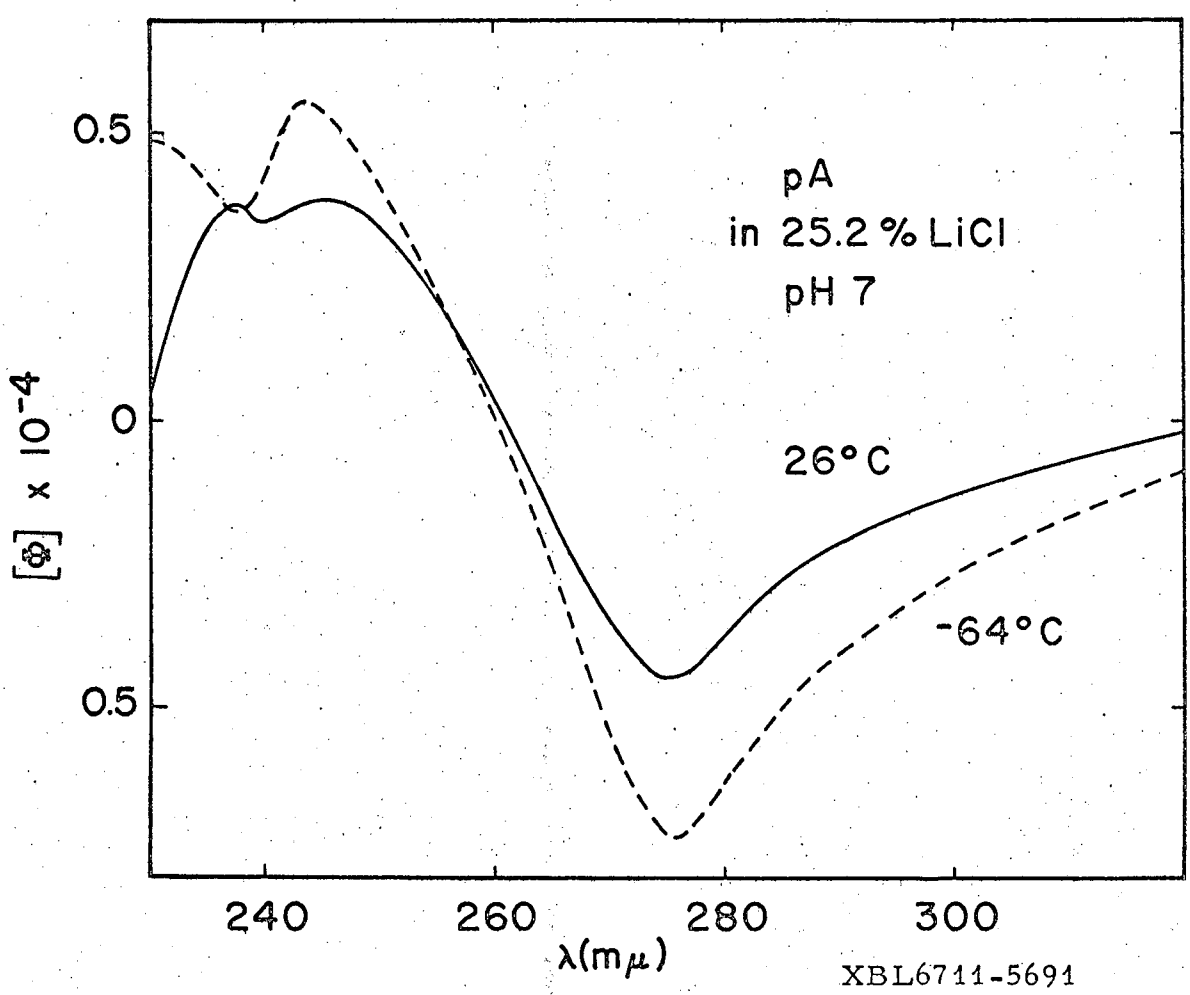



Figure 12 ORD of pA in 25.2% LiCl solution at -64° and 26°C.

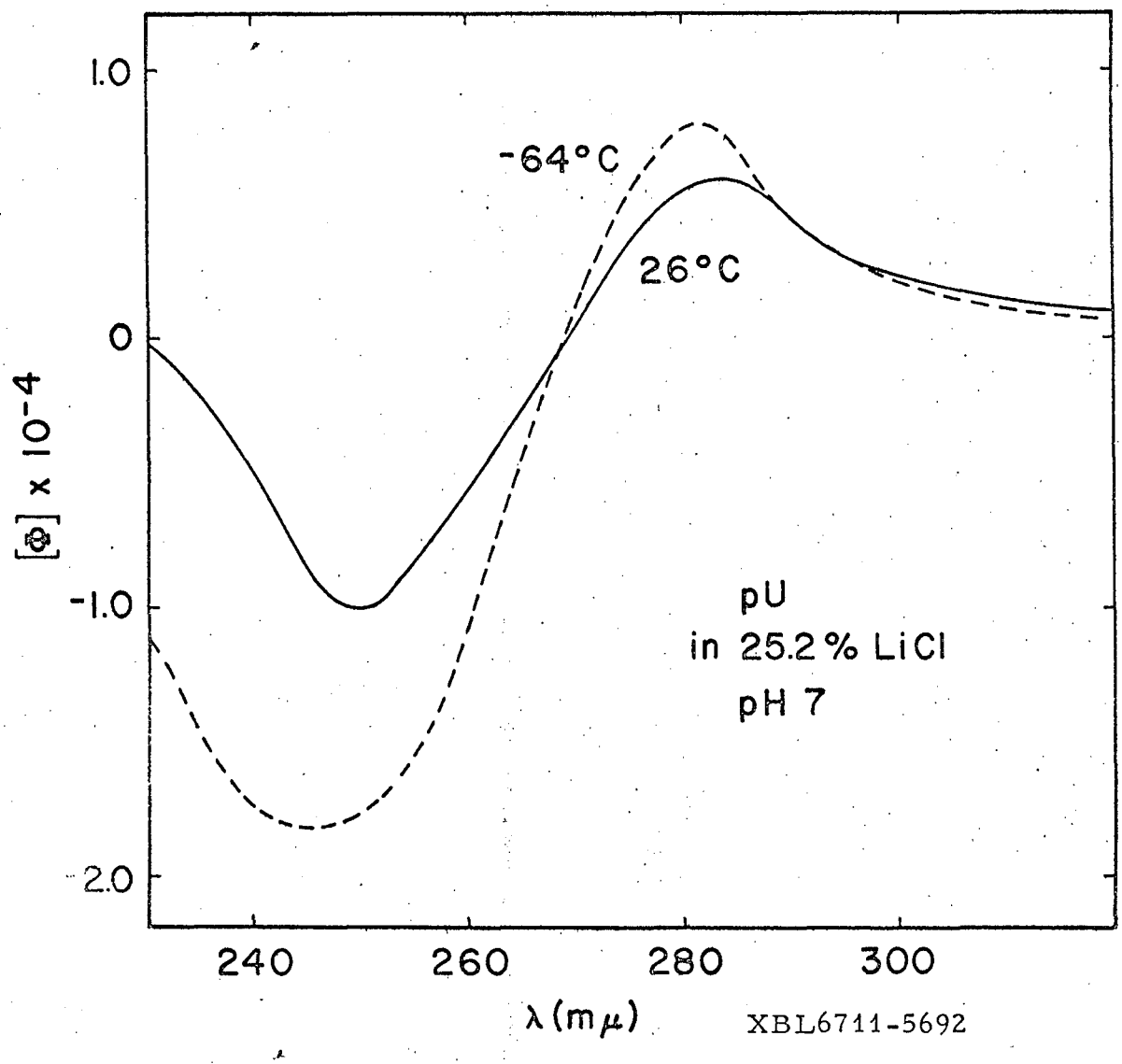



Figure 13 ORD of pU in 25.2% LiCl solution at -64° and 26°C.

its component nucleotides or nucleosides at that temperature.

The low temperature limit is harder to estimate. Theoretical values have been calculated[27], but they are not too dependable in terms of magnitude. In some cases they are seriously in error even in the shape of the ORD. In practice, the value assumed for the low temperature limit was that rotation which gave the most linear plot of log K against 1/T. The assumption of a linear Van't Hoff plot is equivalent to the assumption of nearly equal heat capacities of the "stacked" and "unstacked" dinucleoside phosphates.

Unfortunately evidence to test this assumption is scarce. The temperature dependence of the calorimetric $\Delta H^{o}$ of double-strand formation for poly (A+U) or DNA[24,163] is consistent with this assumption. The temperature dependence of $\Delta H^{o}$ for the double-strand formation of poly (A+U) can be considered as a difference in the heat capacities of the single and double-strand forms. The conformation of the double-strand form is probably independent of temperature. However the single-strand polymers are partially "stacked", with the degree of "stacking" decreasing with increasing temperature. Energy is required for "unstacking". If this is assumed to be the dominant cause of the difference in heat capacities, the temperature dependence of the calorimetric $\Delta H^{o}$ for strand separation can be estimated from the temperature dependence of optical properties. Using the ORD of dinucleoside phosphates as a measure of the amount of "stacking"

and assuming that "stacking" in the polymers is a non-cooperative process one derives an apparent difference of heat capacities for the double and single-strand consistent with the calorimetric data.

The thermodynamic parameters were not very sensitive to moderate changes in the assumed low temperature limit.

TABLE 2

| $[\Phi]_s$ | CpG $\Delta H^o$("unstacking")(in kcal/mole) |
|---|---|
| $1.0 \times 10^4$ | 5.4 |
| $1.5 \times 10^4$ | 4.8 |
| $1.7 \times 10^4$ | 4.8 |
| $2.4 \times 10^4$ | 4.0 |

The wavelength at which the analysis was made is one of three: the first peak, the first trough or the wavelength at which the difference between the rotations of the dinucleoside phosphate and its component nucleosides was a maximum. The wavelength giving the Van't Hoff plot with the least scatter was used. The answer was not dependent on the wavelength chosen, to within experimental error. An estimate of the errors involved in determining the thermodynamic parameters are indicated by the deviation of values for CpG. (See Table 2.)

The fit of the data to the two-state model is reasonably good. (Figures 4-7) The best and worst Van't Hoff plots are shown. (Figure 14) The poorness of fit in the latter case is caused by the difficulties in measuring small rotations.

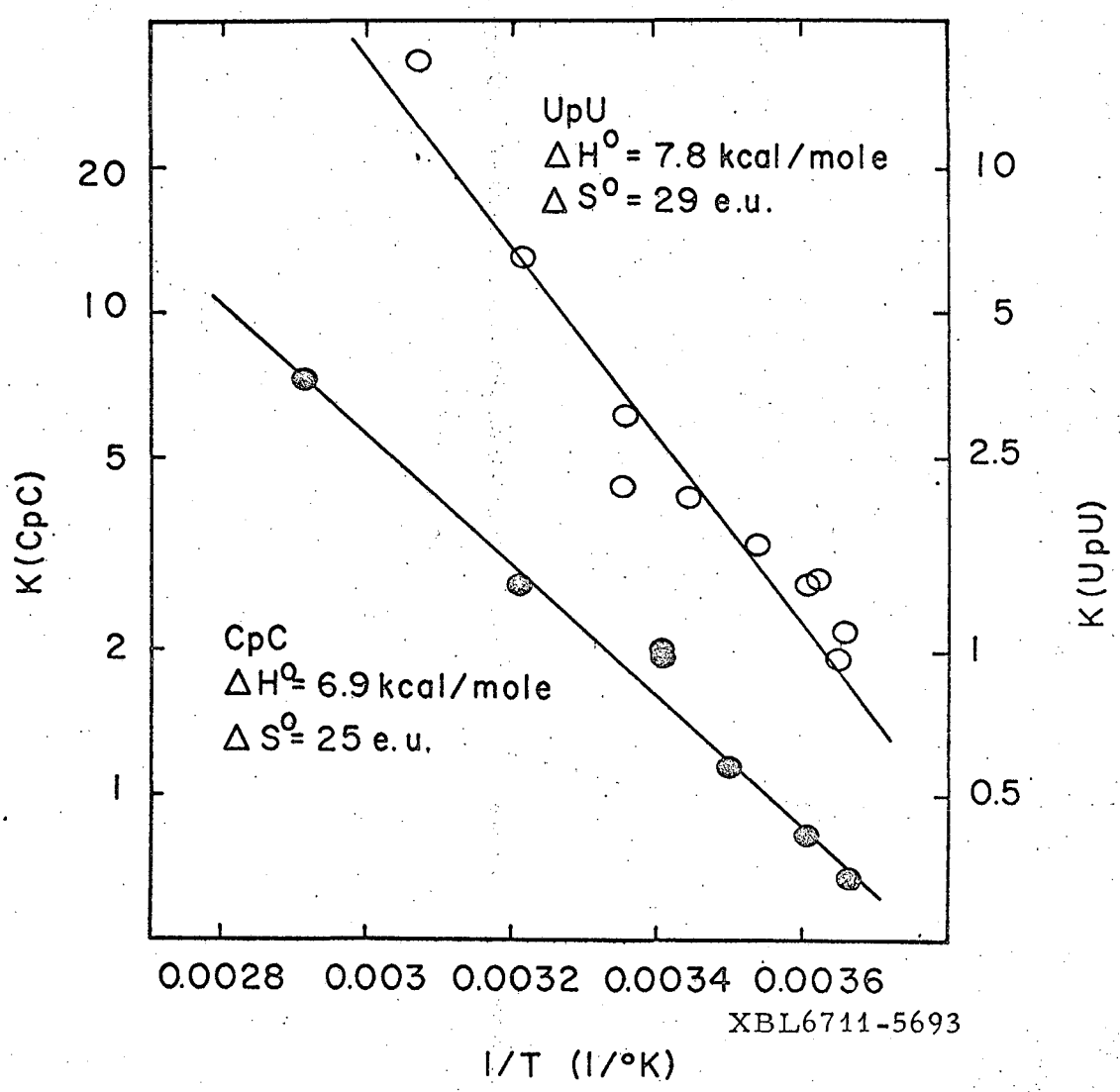



Figure 14 Examples of good and bad fits to the two-state model: Van't Hoff plots for CpC and UpU.

TABLE 3

| Dinucleoside Phosphate | $\Delta H^{o}$ "unstacking" (kcal/mole) | $\Delta S^{o}$ "unstacking" (cal/deg mole) | $\Delta F^{o}$ "unstacking" at 25°C (kcal/mole) | $\kappa$ (cal/mole radian$^2$) |
|---|---|---|---|---|
| ApA | 5.3 | 20 | -0.7 | 200 |
| ApG | 4.8 | 18 | -0.4 | 230 |
| ApC | 6.2 | 22 | +1.7 | 240 |
| ApU | 8.4 | 32 | -1.2 | 340 |
| GpA | 5.6 | 20 | -0.5 | 170 |
| GpC | 7.8 | 28 | +0.6 | 280 |
| GpU | 6.8 | 25 | +0.6 | 140 |
| CpA | 7.3 | 27 | +0.7 | 210 |
| CpG | 4.8 | 17 | -0.2 | 220 |
| CpC | 6.9 | 25 | -0.5 | 300 |
| CpU | 7.8 | 28 | -0.5 | 240 |
| UpA | 5.1 | 21 | -0.7 | 250 |
| UpG | 6.0 | 23 | -0.9 | 130 |
| UpC | 6.2 | 22 | -0.4 | 260 |
| UpU | 7.8 | 29 | -0.9 | 230 |

Even without detailed analysis of these results a few comments should be made. First of all, $\Delta H^o$ and $\Delta S^o$ are all approximately the same (see Table 3) and are comparable with estimates made from the temperature dependence of other optical properties[3,14,15,106,147,203,206] as well as estimates coming from the statistical mechanical analysis of DNA melting profiles[40]. Because different dinucleoside phosphates have similar thermodynamic parameters for the "stacked" to "unstacked" transition it is reasonable to suppose that the origin of "stacking" is not very different in all dinucleoside phosphates. If comparisons are to be made among thermodynamic parameters for the "unstacking" it should be remembered that $\Delta F^o$ is close to zero for all dinucleoside phosphates while $\Delta H^o$ ranges from 4.8 to 8.4 kcal/mole. The percentage uncertainty is greater in the value of $\Delta F^o$ than in $\Delta H^o$.

As a check of the temperature dependence of the ORD of dinucleoside phosphates the temperature dependence of the ultraviolet absorption of some dinucleoside phosphates was measured. Unfortunately the changes in the ultraviolet absorption with changing temperature are small. In most cases the experimental uncertainty was so great as to make any quantitative analysis meaningless. (Figure 15) ApA has one of the largest hypochromic effects measured for a dinucleoside phosphate[211]. The temperature dependence of the oscillator strength (the extinction coefficient integrated on a frequency scale) of ApA is shown in Figure 20.

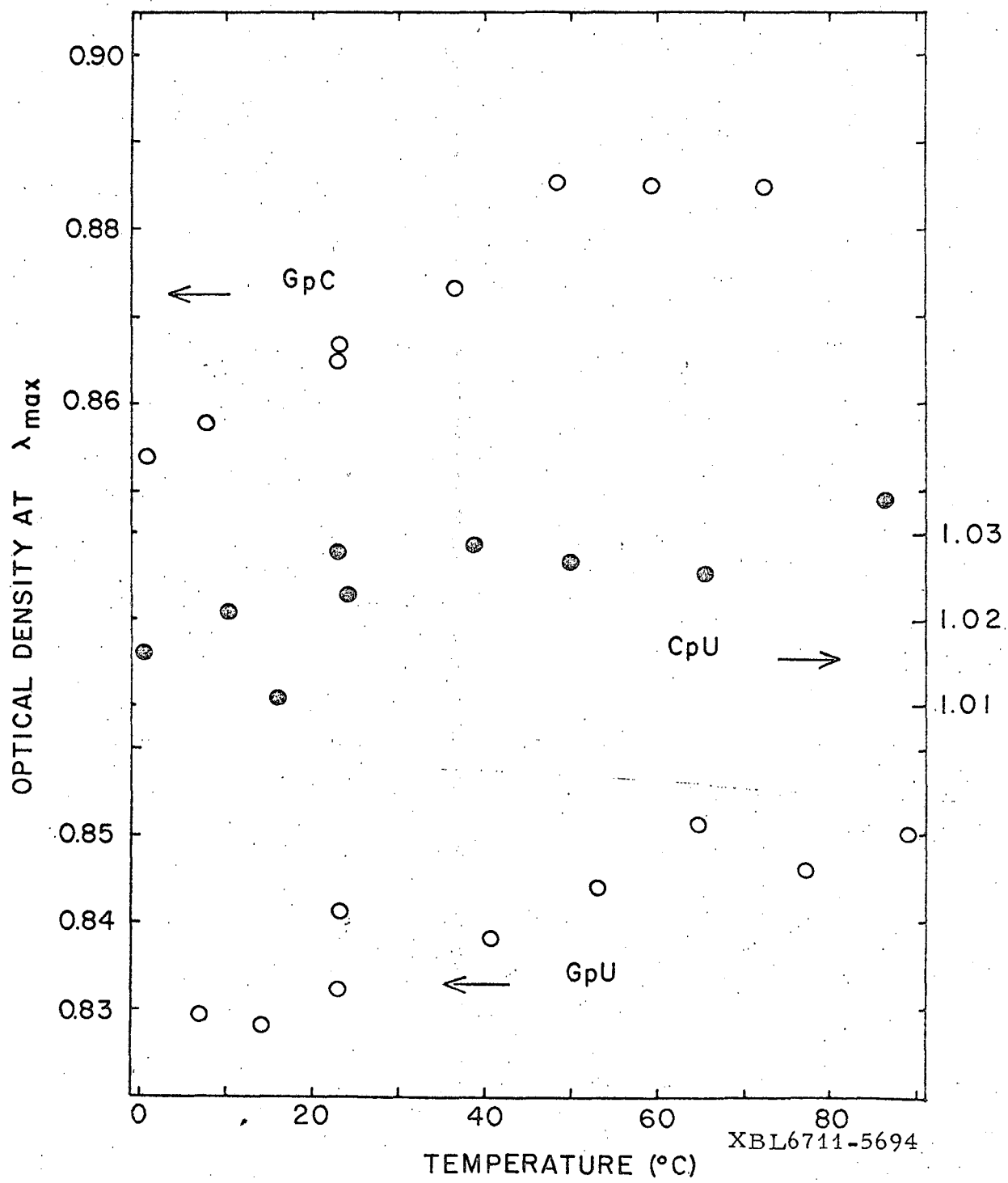



Figure 15 Small changes in absorption of GpC, CpU and GpU with changing temperature.

The change in absorption at the peak is for all intents and purposes the same as the change in the oscillator strength in this case, so that we need only discuss the hypochromism. The hypochromism of ApA can be analyzed in terms of the two-state model. The data fit the model as well as can be expected, considering the experimental scatter. (Figure 16) One fact that should be noted is that the high temperature limit of the oscillator strength is not that of adenosine. This agrees with other measurements of the hypochromicity of ApA[106]. Considering this in terms of the two-state model, the "unstacked" conformation would still have appreciable hypochromism. This suggests that even at high temperatures the bases are close enough to one another to permit appreciable hypochromism. Unfortunately, since a reliable quantitative theory for the hypochromic effect does not exist we can not estimate whether a dinucleoside phosphate having bases 5 or 10 Å away from one another (i.e. "unstacked") might have 3% hypochromism.

$\Delta H^o$ ("unstacking") determined from this analysis is 8.5 kcal/mole (Figure 16). This is within the range of values of $\Delta H^o$ ("unstacking") obtained by analyzing the temperature dependence of the ORD of dinucleoside phosphates. However it is considerably larger than the value of 5.3 kcal/mole determined for ApA. This fact makes the model itself suspect. If "unstacking" is indeed a two-state process then one should calculate the same values for the thermodynamic parameters using any method which is capable of seeing the

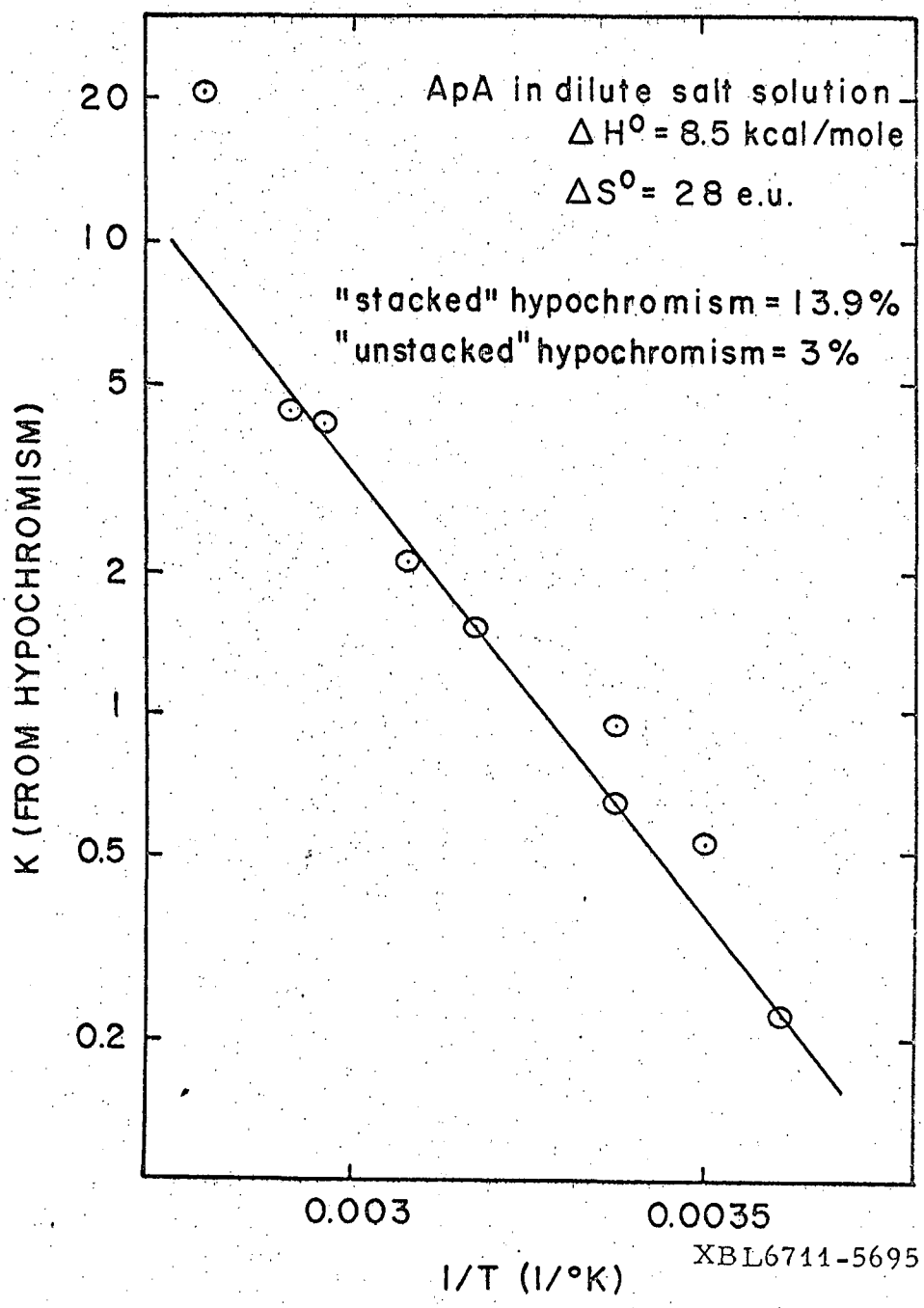



Figure 16 Van't Hoff plot for ApA in dilute salt solutions using hypochromism as a measure of "stacking".

difference between the "stacked" and "unstacked" forms. For ApA this is not the case; values of $\Delta H^{o}$ range from 5.3 to 10 kcal/mole. Some of the difference may be caused by differences in methods of analysis. The choice of a low temperature and high temperature limit varies from one analysis to another. Even considering this, it is hard to reconcile a two-fold variation in a parameter which should be a constant.

TABLE 4

$\Delta H^{o}$("unstacking") for ApA

| Observable | $\Delta H^{o}$("unstacking") | Reference |
|---|---|---|
| ORD | 5.3 kcal/mole | This work |
| ORD | 6.5 kcal/mole* | 147 |
| CD | 8 kcal/mole | 203 |
| Hypochromism | 8.5 kcal/mole | This work |
| Hypochromicity<br>Calorimetry (poly A) | 9 kcal/mole | 56 |
| Hypochromicity | 9.4 kcal/mole | 3 |
| Hypochromicity | 10 kcal/mole | 106 |

* $\Delta H^{o}$=6.5 kcal/mole was determined using data from a series of oligoadenylic acids. If the temperature dependence of only ApAp is considered the best value for $\Delta H^{o}$ is 6.1 kcal/mole.

Another objection to this model is that it is too simple. In considering a system such as double-strand DNA, it is reasonable to assume a two-state model. One expects a unique ordered form different from the disordered form, with a sharp transition between them. However, in

the case of dinucleoside phosphates there is probably a range of "stacked" conformations which together with the "unstacked" conformations form a continuous spectrum. There is no reason to expect a sharp division on the molecular level between the two regions; this is an assumption basic to the two-state model.

There has been a treatment comparing two-state and multi-state analyses of just such processes[117]. Two tests are proposed to check the validity of the two-state assumption. One is the independence of $\Delta H^o$ of the technique used to measure the transition. In the case of the "unstacking" of dinucleoside phosphates there is some variation. We must consider how much of this variation results from the analyses of the data. ORD and CD (as analyzed here and in the literature) should give identical answers. They do not. The range of all values for $\Delta H^o$ ("unstacking") for ApA is not too much larger than the difference between the values obtained from CD and ORD measurements. There is a chance that all deviations of $\Delta H^o$ are caused by problems in measurement and analysis, rather than the "unstacking" process itself. This is quite surprising considering that the estimates of errors in these numbers in our work is far smaller than the reported variations.

The second test proposed depends on the variation of $\Delta H^o$ with temperature. In the case of dinucleoside phosphates the data are analyzed in such a way to give a constant $\Delta H^o$ (linear Van't Hoff plot). Therefore this test can not be applied.

Although the evidence against the validity of the two-state model is incomplete at this point, the meaning of the thermodynamic parameters derived from this treatment should be considered in light of the possibility that the two-state model may be wrong. The apparent $\Delta H^o$ (as determined by two-state analysis) can be badly in error in such situations[117]. We do have evidence that this is probably not the case here. The calorimetric[66,179] and statistical mechanical[40] studies already mentioned do give $\Delta H^o$ ("unstacking") in the same range as ours. Although the systems studied were not dinucleoside phosphates they were similar enough to add credulity to the values determined by the two-state analysis. Considering all the evidence available we must conclude that the thermodynamic parameters determined from measurement of the temperature dependence of the ORD (as a measure of the average conformation) using a two-state method of analysis are reasonable. We expect that the difference in enthalpy between the high and low temperature limits of dinucleoside phosphates be between 4 and 10 kcal/mole. However, this does not insure that comparisons of $\Delta H^o$ and $\Delta S^o$ for the disordering process among various dinucleoside phosphates are meaningful.

## OSCILLATING DIMER MODEL

The inconsistencies of the two-state analysis suggest that another approach to the problem is called for. The oscillating dimer model, although overly simple, has two attractive features, a continuum of allowed conformations

and different temperature dependence for different physical properties.

Since the temperature dependence of the ORD has been explicitly calculated (using the dipole-dipole approximation) for the case where the bases remain parallel to one another and directly above one another while undergoing torsional oscillations, quantitative comparisons with the data are possible. This model predicts that in the classical (high temperature) limit log [Φ] should vary linearly with temperature, with the slope being inversely proportional to the torsional force constant. This method of analysis has an advantage over the two-state model. The low temperature limit does not have to be known in order to analyze the data. The temperature dependence of the ORD of the dinucleoside phosphates fit this model as well as they do the two-state model. The best and worst fits of the data to this model are shown in Figure 17. As was the case with the Van't Hoff plots, much of the larger deviation in the latter case is caused by difficulties in measuring very small rotations.

Results of the analysis of the data using this model are given in Table 3. The apparent force constants are approximately one or two orders of magnitude smaller than for molecules such as chloro-substituted ethanes[112]. This is to be expected because one is attributing all the motion of one base relative to the other to a single rotational degree of freedom. In reality, there are seven covalent

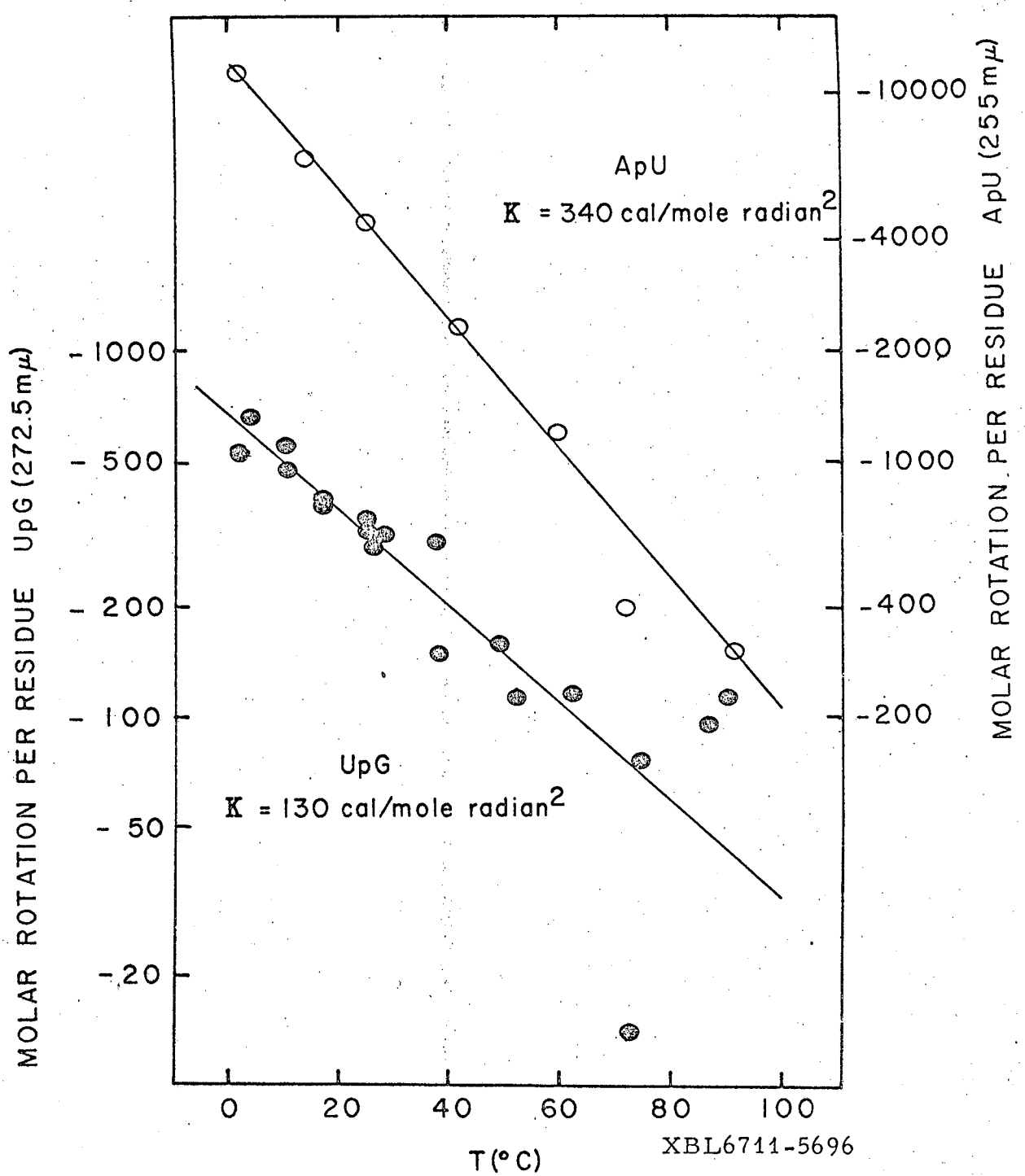



Figure 17 Examples of good and bad fits of the ORD to the oscillating dimer model: ApU and UpG.

bonds which can be rotated more or less freely. Small oscillations about a number of these bonds would show up as large fluctuations in the single rotational degree of freedom, making the apparent force constant smaller than any single force constant.

LOW TEMPERATURE MEASUREMENTS

The fact that the experimental data available from dinucleoside phosphates in dilute salt solutions between 0°C and 85°C fit two very different and very crude models equally well made it important to make further measurements which might enable us to say more about the conformations of dinucleoside phosphates in solution. Use of concentrated solutions of LiCl in $H_2O$ enables us to greatly extend the temperature range over which we could measure optical properties; 25.2% LiCl in $H_2O$ freezes at -75.7°C. This allows us to measure experimentally the low temperature limit of a physical property instead of having to assume one. Furthermore, it nearly doubles the temperature range which can be used to test a model. Since 25.2% LiCl corresponds to one LiCl per seven $H_2O$ it is necessary to determine whether dinucleoside phosphates behave in this solvent as they do in dilute salt solutions.

One specific possible difficulty arising because of the use of concentrated LiCl solutions can be ruled out. $Li^+$, having a very high charge to diameter ratio, resembles $H^+$. At these high concentrations $Li^+$ might interact strongly

with non-bonded electrons of the bases just as protons do. This is not the case. The ultraviolet absorption of nucleotides change greatly on protonation[205]. The shape of the 260 mμ absorption band of adenylic acid changes, becoming quite skewed. In the case of cytidylic acid there are two strong absorption bands at neutral pH and only one strong one below the cytosine pK. The absorption band shapes of adenylic acid and cytidylic acid do not change greatly in going from dilute aqueous buffers to 25.2% LiCl. There are slight shifts towards the red, which is expected when a $\pi$-$\pi^*$ transition is exposed to a more polarizable solvent[125], but nothing indicating strong specific interactions with the solvent.

A less direct piece of evidence comes from the study of infrared absorption of concentrated salt solutions[74]. The shift of frequency of the OH stretch in LiCl solutions even more concentrated than those used in these studies is surprisingly small. There is a shift of about 10 $cm^{-1}$ in going from pure water to 2 M LiCl, with essentially no additional shift for more concentrated solutions. This is compatible with the idea that the microscopic structure of solutions of 25.2% LiCl in $H_2O$ is quite similar to that found in more dilute solutions.

In certain cases similarities in the properties of dinucleoside phosphates in dilute buffers and 25.2% LiCl are striking. Using the two-state model to analyze the temperature profile of the ORD of ApA in 25.2% LiCl, the

same values of the thermodynamic variables were obtained as for the dilute salt solutions. (Figure 18) The actual ORD of ApA in 25.2% LiCl is very much like that in dilute salt solutions. There are only two major differences. The magnitude of the rotation is 25.2% LiCl is smaller by about 35% and the ORD (and ultraviolet absorption as well) is shifted to the red by a few mμ. However, there are differences in the thermodynamic parameters for the "unstacking" of the other dinucleoside phosphates measured. $\Delta H^{o}$ ("unstacking") is smaller in 25.2% LiCl than in dilute buffer for CpC (4.9 compared to 6.9 kcal/mole) and UpC (5.0 compared to 7.8 kcal/mole). For UpU in 25.2% LiCl there is no measurable interaction between the bases contributing to the ORD above 0°C. In dilute salt solutions the bases do seem to interact slightly[210]. In general we think that dinucleoside phosphates in concentrated LiCl solutions will behave similarly to those in dilute salt solutions.

The temperature dependence of the ORD of ApA in the concentrated LiCl solutions is even larger than the change in dilute buffers. (Figure 18) Careful examination of the data provides an opportunity to critically check the two-state and oscillating dimer models. The dependence of the rotation at the first trough (262 mμ) is shown in Figure 18. The fit to the two-state model seems good. The Van't Hoff plot of the data (Figure 19) allows closer examination. There seems to be hints of a systematic deviation from the predictions of two-state model, a slight sinusoidal type

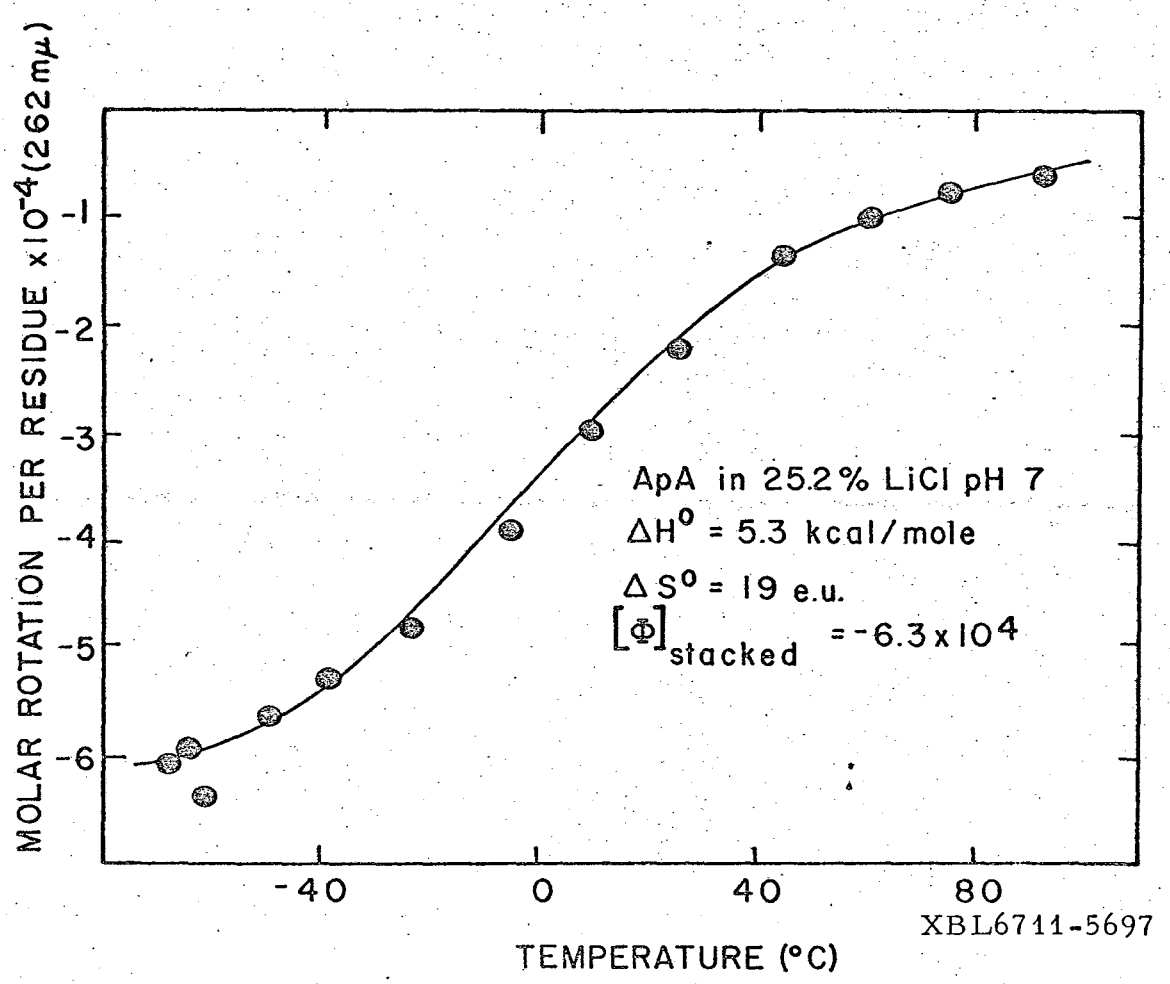



Figure 18 $[\Phi]_{262}$ of ApA in 25.2% LiCl solution as a function of temperature. The curve drawn through the points is calculated using the two-state model.

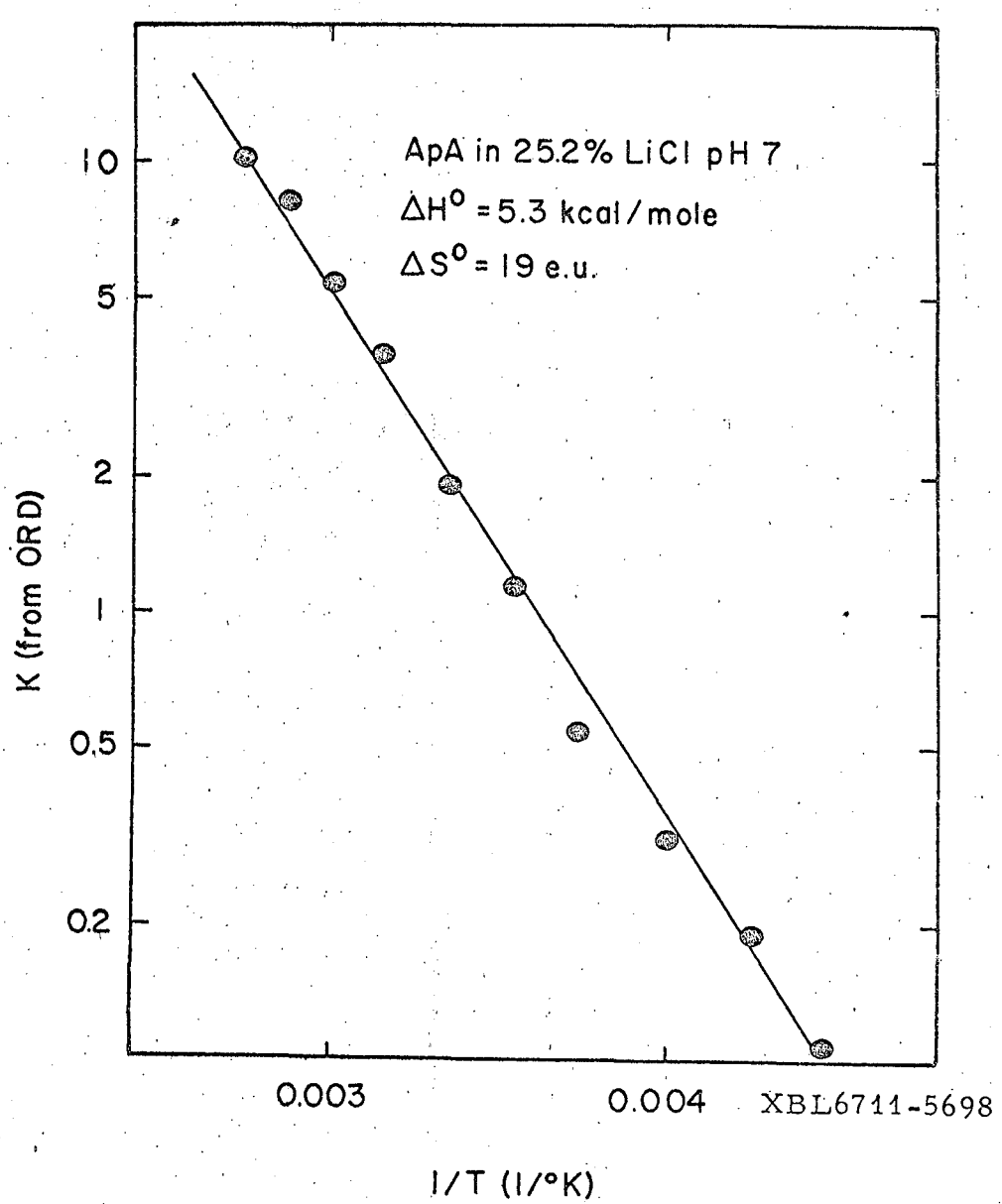



Figure 19 Van't Hoff plot for ApA in 25.2% LiCl solution using ORD as a measure of "stacking".

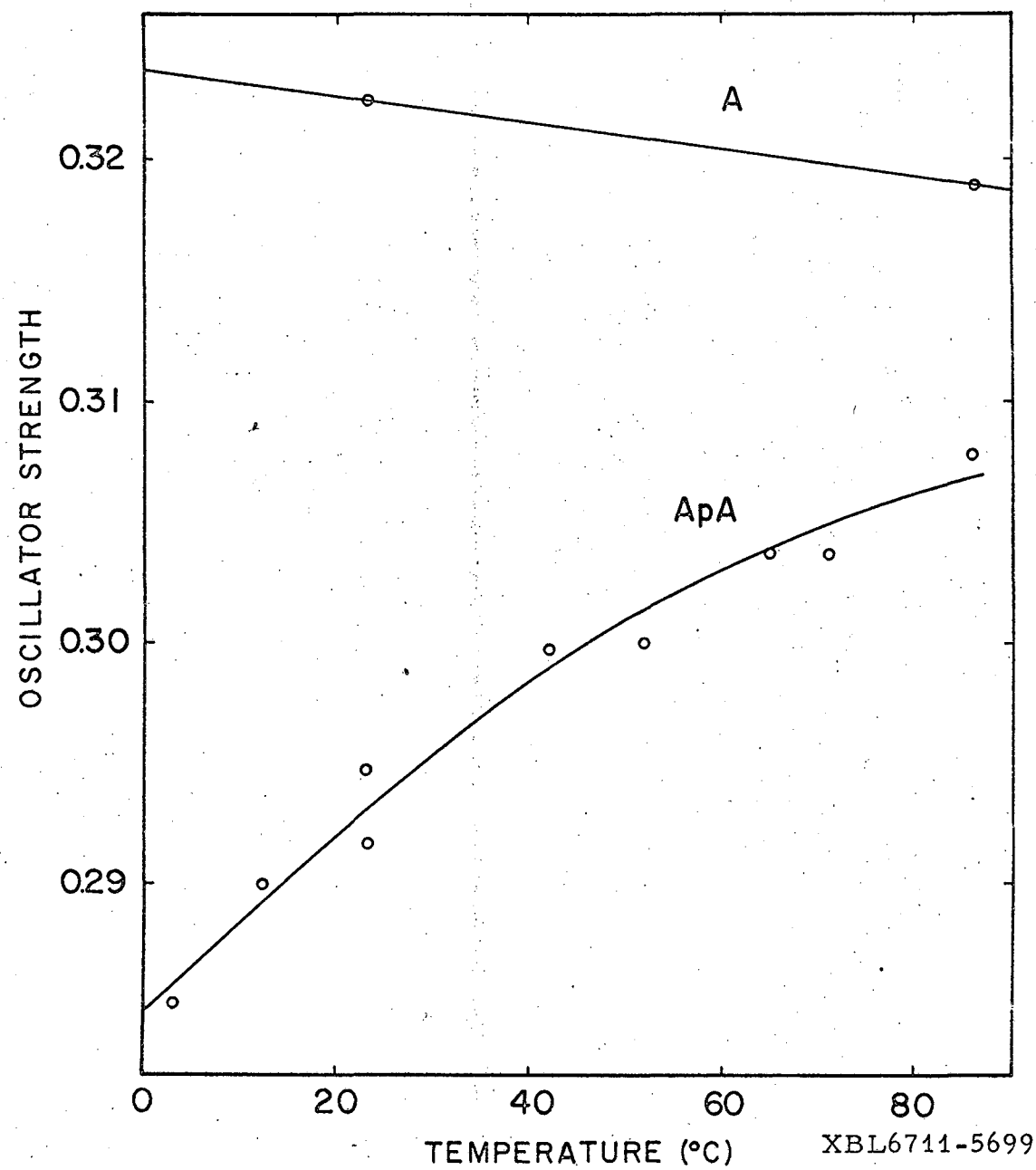



Figure 20 Oscillator strength of ApA and pA as a function of temperature.

of deviation from a straight line. Unfortunately the data is not precise enough to have much confidence in such deviations. This illustrates the problems in choosing between models. Quite different models may have very similar temperature dependences of a physical property. We will see later that there is unambiguous evidence that the two-state model does not fit the data.

The classical limit of the oscillating dimer model (which predicts an exponential increase of the ORD with decreasing temperature) can not fit a curve which starts to level off at low temperatures. Since the low temperature limit is where quantum effects become most important we should investigate the possibility of the quantum mechanical expression's fitting the data. It fails on two points. The levelling off of the ORD of ApA occurs about 220°K to 270°K. This is the region in which $\hbar\omega$ should be approximately equal to kT. Expressing this in terms of the moment of inertia of the base (using the force constant already determined); this leads to a value of about $10^{-41}$ $gcm^2$ for the moment of inertia of a disc representing a base. This is approximately two orders of magnitude smaller than the moment of inertia of a base around any conceiveable axis. In addition, the levelling off of the ORD of ApA is much more sudden than the predictions of the quantum mechanical oscillating dimer model using the very small moment of inertia. These are manifestations of the excessive simplicity of the oscillating dimer model.

Although the temperature dependence of the ORD of ApA in 25.2% LiCl (Figure 18) can be fit by a two-state model, the analogous data for the absorption (Figure 2) of the same solution can not. The complete lack of agreement between the ORD and absorption temperature dependences shows clearly that a two-state model for a dinucleoside phosphate is inadequate. Each physical property has a different geometric dependence and therefore a different temperature dependence. This may explain the large differences in apparent $\Delta H^{o}$ values obtained from different measurements (see Table 4).

The absorption data give more specific information about dinucleoside phosphate structure. At -67°C the ultraviolet absorption spectrum shows vibronic structure similar to that seen for adenine in non-polar matrices at liquid helium temperature[192]. (Figure 21) Although the absorption of pA in 25.2% LiCl sharpens slightly at -67°C it does not show as much structure as is seen for ApA. This is caused by the stacking of bases in ApA. Each base shields the other from the polar solvent, making the environment of the chromophore look less polar than in pA where each base is essentially surrounded by the polar solvent.

The fact that the absorption of ApA is higher than pA at temperatures higher than 90°C is striking. It can be explained in terms of solvent induced hypochromism. As the concentrations used in these experiments (greater

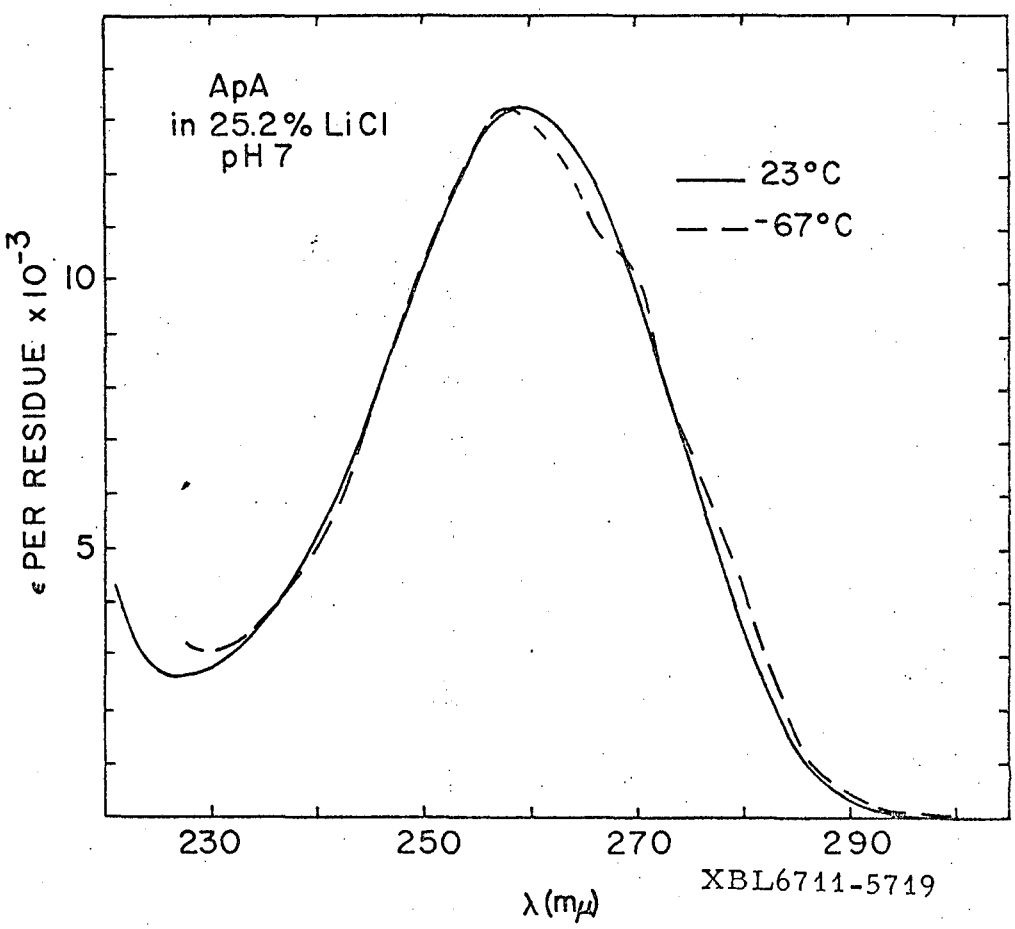



Figure 21 Ultraviolet absorption spectra of ApA at -68° and 25°C.

than 7M) $Cl^-$ absorbs light very strongly at 200 mμ. The dispersion of the polarizability resulting from this absorption gives rise to a large polarizability at 260 mμ. The 260 mμ transition of adenine could interact with this polarizability giving the solvent hypochromism observed. (The oscillator strength of the 260 mμ absorption of pA decreases from .322 in dilute salt buffers to .311 in 25.2% LiCl.) At low temperature the base-base interaction dominates the hypochromism. At high temperatures the average distance between the centers of the bases increases and the hypochromism caused by base-base interaction becomes smaller than the hypochromism caused by the highly polarizable solvent. Because the base in pA is fully exposed to the solvent while the bases of ApA partially shield one another even at high temperatures, pA is more hypochromic than ApA. The bases oscillate with respect to one another (pivoted at the sugar-phosphate bonds), but remain parallel with no solvent between them.

This is compatible with the temperature dependence of the hypochromism of ApA in dilute slat buffers. In that case the high temperature limit of the oscillator strength of ApA is approximately 3% less than that of pA, indicating little solvent induced hypochromism. At high temperatures the bases are still close enough to one another so that their interaction gives hypochromism of this magnitude.

The temperature dependence of the ORD and ultraviolet hypochromism suggest a continuous model for a dinucleoside phosphate. The temperature dependence of the ORD of ApA

in 25.2% LiCl fits the oscillating dimer model at high temperatures, but at low temperatures it does not. For non-rigid, dynamic models the shape of the potential energy surface determines the temperature dependence of an observable property, so it is not surprising that the one dimensional harmonic oscillator does not give the correct shape. (A one dimensional cosine function gives an even poorer fit[116].) The measured temperature dependences of various different properties should provide a good criterion and test for more detailed potential functions.

NUCLEAR MAGNETIC RESONANCE MEASUREMENTS

Further evidence for the structure of a dinucleoside phosphate comes from NMR measurements of ApA, A, pA and Ap. Each adenine base of ApA dissolved in $D_2O$ containes two protons: $H_2$ and $H_8$. If the corresponding protons on the two bases are equivalent there will be only two resonances in the aromatic region of the NMR spectrum. Measurements of ApA at three different temperatures showed four peaks in the aromatic region assigned to the four magnetically non-equivalent protons (see Figure 22). These four peaks were still distinct even at 90°C. Two effects could cause the non-equivalence: intrinsic non-equivalence caused by the phosphate ester at the 3' and 5' ribose positions, or non-equivalence arising from interaction with the other base. The possibilities were tested in two ways. Measurements of 5' adenylic acid (pA), mixed 2', 3' adenylic acid (Ap) and adenosine (A) were made. The NMR signals of the adenine $H_2$ and $H_8$ of A are identical to those of pA. This

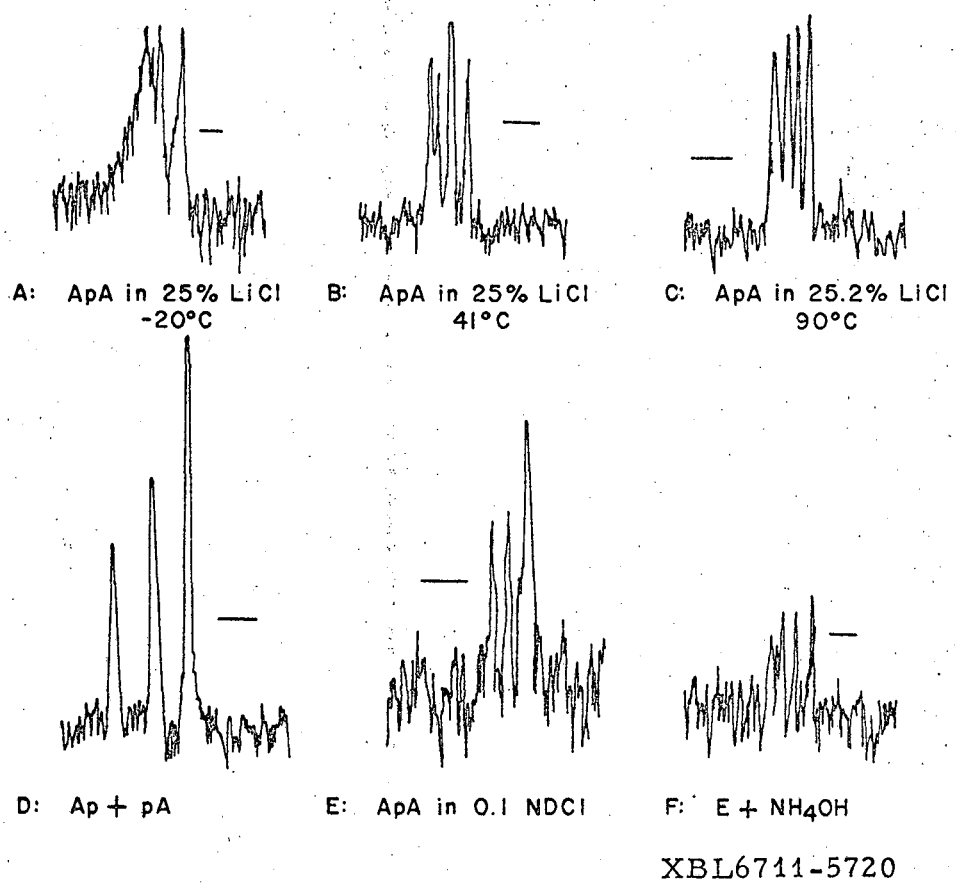



Figure 22 NMR spectra of the aromatic regions of ApA, Ap and pA. The horizontal lines indicate 10 cycles per second.

is not surprising because the 5' phosphate is quite far from the base. The NMR signal of the $H_2$ of Ap was the same as the others, but the $H_8$ proton signal was shifted. These results agree with previous observations[87]. Again, the distance of the $H_2$ from the phosphate of Ap is considerably greater than the distance of the $H_8$ from the phosphate. A smaller effect is expected for the proton farther away. These results lead to the conclusion that the intrinsic non-equivalence of Ap and pA can at most lead to three lines. Measurement of a mixture of these monomers gave a three line spectrum (see Figure 22) distinct from the four line spectrum of ApA. A more direct test comes from the NMR of ApA in acid solution where coulombic repulsion of the positively charged bases destroys the ordered "stacked" structure[210]. The NMR of ApA in 0.1 N HCl shows three resonances in the aromatic region (Figure 22), quite similar to the mixture of the isomers of adenylic acid. After the spectrum was taken, the solution was made basic with concentrated $NH_4OH$. This solution gave a four line spectrum in the aromatic region, indicating that no appreciable hydrolysis or depurination had taken place. Unfortunately the interpretation of the data is not unambiguous. Protonation reduces the aromatic "ring current"[198] as well as the "stacking" of bases[210]. It is difficult to say whether the three line NMR spectrum in the aromatic region of protonated ApA is caused by "unstacking" of ApA alone, or by reduction of the perturbation of one base on the protons of the other. However the NMR of unprotonated ApA in dimethyl sulfoxide (an "unstacking" solvent) has been measured and found to be

almost identical to the spectrum of the mixed 3' and 5' nucleotides[198]. We can therefore say that the non-equivalence of the high field aromatic protons in ApA at 90°C is caused by base-base interactions (which are comparable in size to those at room temperature) rather than the differences in the phosphate ester linkages. (Professor Sunney Chan, California Institute of Technology, has pointed out that intermolecular base-base interactions can also cause magnetic non-equivalence. These effects may also be present in our solutions particularly at the lowest temperature. Of course this does not change the conclusion that strong intramolecular base-base interaction occurs even at the highest temperature.)

These experiments lead to the conclusion that ApA (and probably other dinucleoside phosphates) in solution maintain conformations which hold the two bases close to one another with no "unstacked" dinucleoside phosphate (where only covalent bonds limit the motion of the bases with respect to one another) probably does not exist for ApA in aqueous solutions under experimental conditions used until now.

Measurements of the methyl proton NMR of thymine in various denatured DNAs also indicate that "stacking" remains at high temperatures[124]. Two peaks are found for these methyl protons in DNA above $T_m$. The intensities of these two peaks correlate well with the probability of finding a purine base attached to the 5' phosphate of

thymine. (These nearest neighbor distributions had already been determined by chemical methods[91].) The thymine methyl group is almost directly below the base of its 5' neighboring nucleotide in a right handed "stacked" configuration, while it is quite far from its 3' neighbor. Because purine and pyrimidine bases have quite different "ring currents" two methyl peaks are seen, one caused by interaction with the adjacent 5' purine nucleotide, the other by interaction with the adjacent 5' pyrimidine nucleotide. These measurements were made at and above 90°C, showing that there is still appreciable "stacking" in DNAs at high temperatures. This strongly supports our finding that there is considerable base-base interaction in ApA at these temperatures.

COMPARISONS AMONG DINUCLEOSIDE PHOSPHATES

Although we have discussed the properties of only ApA in concentrated LiCl solution down to lower temperatures we feel that it is representative of a number of dinucleoside phosphates. The temperature dependence of the ORD of CpC (Figure 23, 24) is quite similar to that of ApA. The data also fits the two-state model well, with a midpoint of the "stacking" process at about 0°C. We feel that this pattern will also be found for other properties and other dinucleoside phosphates. In these cases the bases of the dinucleoside phosphates remain close to one another throughout the entire range of temperature experimentally accessible.

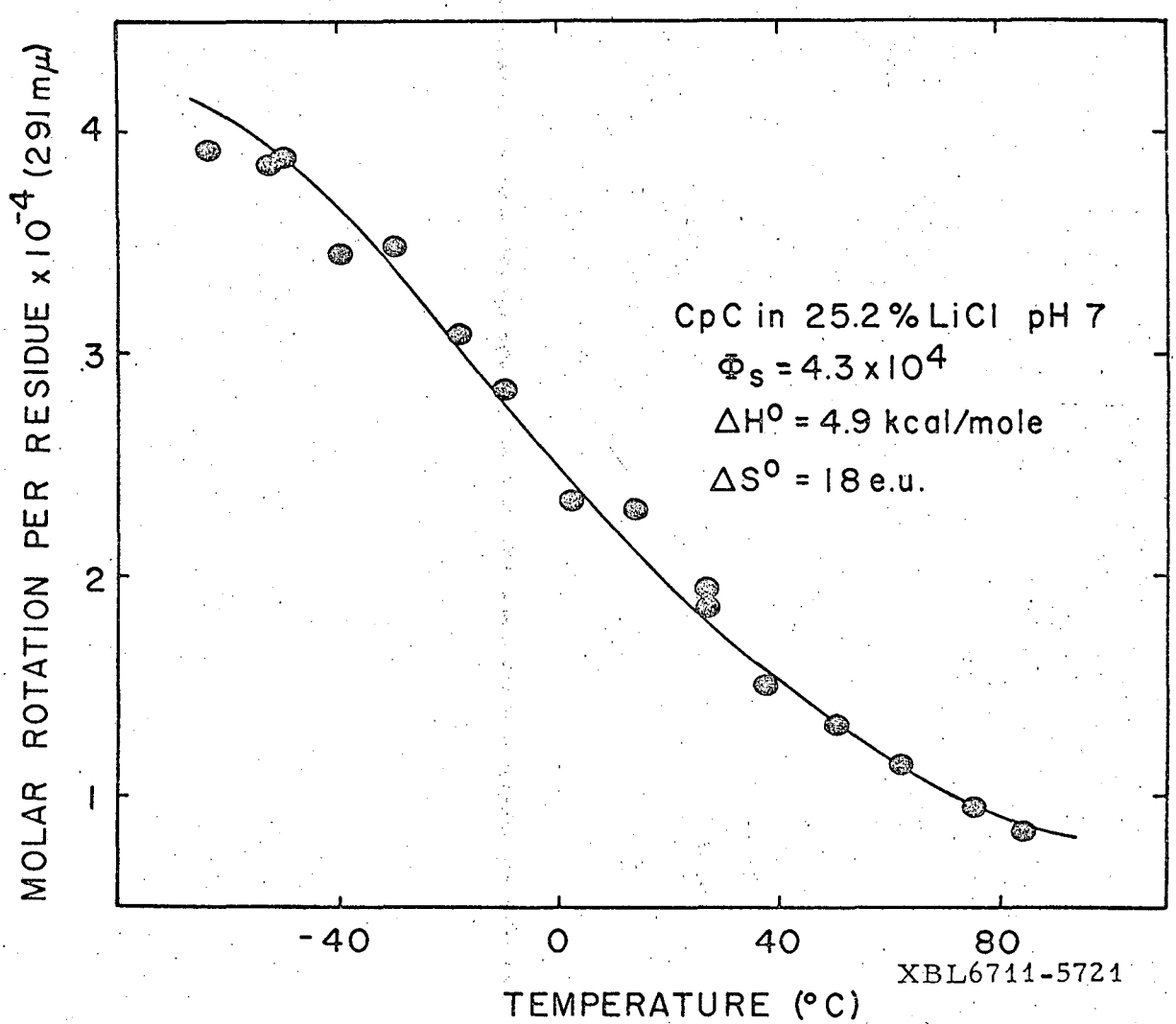



Figure 23 ORD of CpC in 25.2% LiCl solution at -64°, 26.5°, and 84°C.

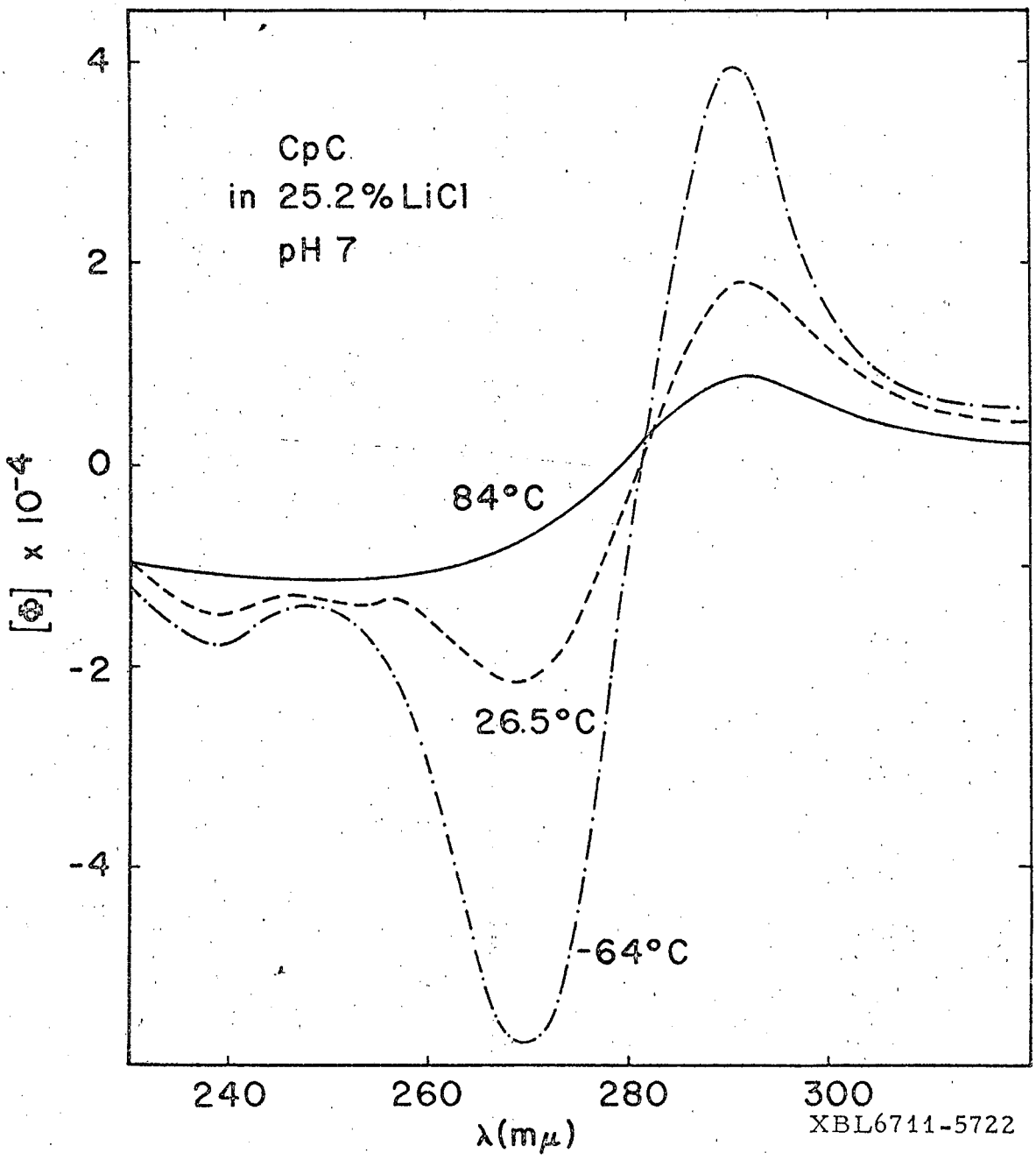



Figure 24 $[\Phi]_{291}$ of CpC as a function of temperature. The curve drawn through the points is calculated using the two-state model.

Since the helix to coil transition in single-strand polynucleotides is essentially non-cooperative[3,16,106,147] we expect the base-base interactions in single-strand polynucleotides to be similar to those in dinucleoside phosphates. Therefore, a single-strand polynucleotide at the mid-point of its helix-coil transition would tend to have all of its bases half "stacked" rather than half of its bases fully "stacked" and half fully "unstacked". This means that there might be a considerable amount of short range disorder while still retaining a fair amount of long range order. In aqueous solutions single-strand polynucleotides such as polyadenylic acid or polycytidylic acid should be considered wormlike coils rather than a series of stiff rods connected by universal joints. Low angle X-ray diffraction studies of concentrated polyadenylic acid solutions at neutral pH show that the molecules appear rod-like over very long distances with an electron density corresponding to one nucleotide per 3.4Å[118]. There are two interpretations suggested by the authors. Polyadenylic acid could be either a single-strand highly "stacked" structure or a double-strand intercalated structure. Although polyadenylic acid at neutral pH in dilute solutions has been shown to have a single-strand structure[79] we prefer the latter interpretation of the low angle X-ray scattering experiments. The gels used in these studies were very much more concentrated than the solutions used in the studies of optical properties. In addition, the rigidity of the rods of polyadenylic acid are much larger than expected,

for single-strand polynucleotides, being considerably more rigid than the neutral form of polyadenylic acid studied in hydrodynamic experiments[79] and comparable or greater in rigidity compared with native DNA[69,199]. It is much easier to see how a double-strand species might be this rigid. Unfortunately we have little information about the multi-dimensional energy surface of ApA. This information (or equivalent information about the distribution of conformations) is necessary in order to estimate the statistics of the orientation of neighboring nucleotides and the rigidity of polyadenylic acid in neutral aqueous solutions.

The continuous nature of the potential energy of a dinucleoside phosphate (as opposed to a small number of allowed conformations) suggests a dynamic structure for dinucleoside phosphates. Since there should not be extensive high energy barriers there would be no reason for a molecule to stay in a single conformation for an extended period of time. Tritium exchange studies have demonstrated the dynamic nature of DNA and transfer RNA on a relatively long time scale[55,148]. The motion we expect is very much faster. NMR measurements are compatible with this type of motion[33]. The line broadening of proton resonances of purine when interelated between the bases of a dinucleotide is explained in terms of enhanced nuclear spin relaxation of those protons which is induced by the fluctuating magnetic field they see. This magnetic field is caused by the magnetic moments of the nuclei of the dinucleotide. The fluctuations result

from a motion which has a characteristic time of approximately $10^{-10}$ seconds. Although this motion may be caused by the purine molecule "rattling" in its cage, it may also be caused by a rapid motion which is inherent in the dinucleotide. This is just the type of motion we should expect as a result of the continuous nature of the allowed conformations of dinucleoside phosphates. Rapid motion of this type may be another manifestation of short range disorder without complete loss of long range order. This might give rise to large differences in the behavior of properties which depend on the degree of short range order (ORD, ultraviolet absorption, NMR) and those which depend on the degree of long range order (sedimentation velocity, viscosity).

The ORD of ApA and CpC at 25°C are considerably different from their constituent nucleosides and nucleotides. The ORD of dinucleoside phosphates such as UpU or UpC are not. There are two possible explanations. Either the bases are not close to one another, are less "stacked" than ApA, for example[210,211] or the optical properties of fully "stacked" UpU and UpC are just not very different from their constituent nucleoside and nucleotides[127]. (Large differences in the magnitudes of the ORD of different "stacked" dinucleoside phosphates have been predicted on the basis of quantum mechanical calculations[27].) By going to lower temperatures a more reliable value for the rotation of the fully "stacked" dinucleoside phosphate can be determined and this dilemma, resolved.

The ORD of two dinucleoside phosphates in this category were measured down to -64°C. (Figures 25-27) The inter-

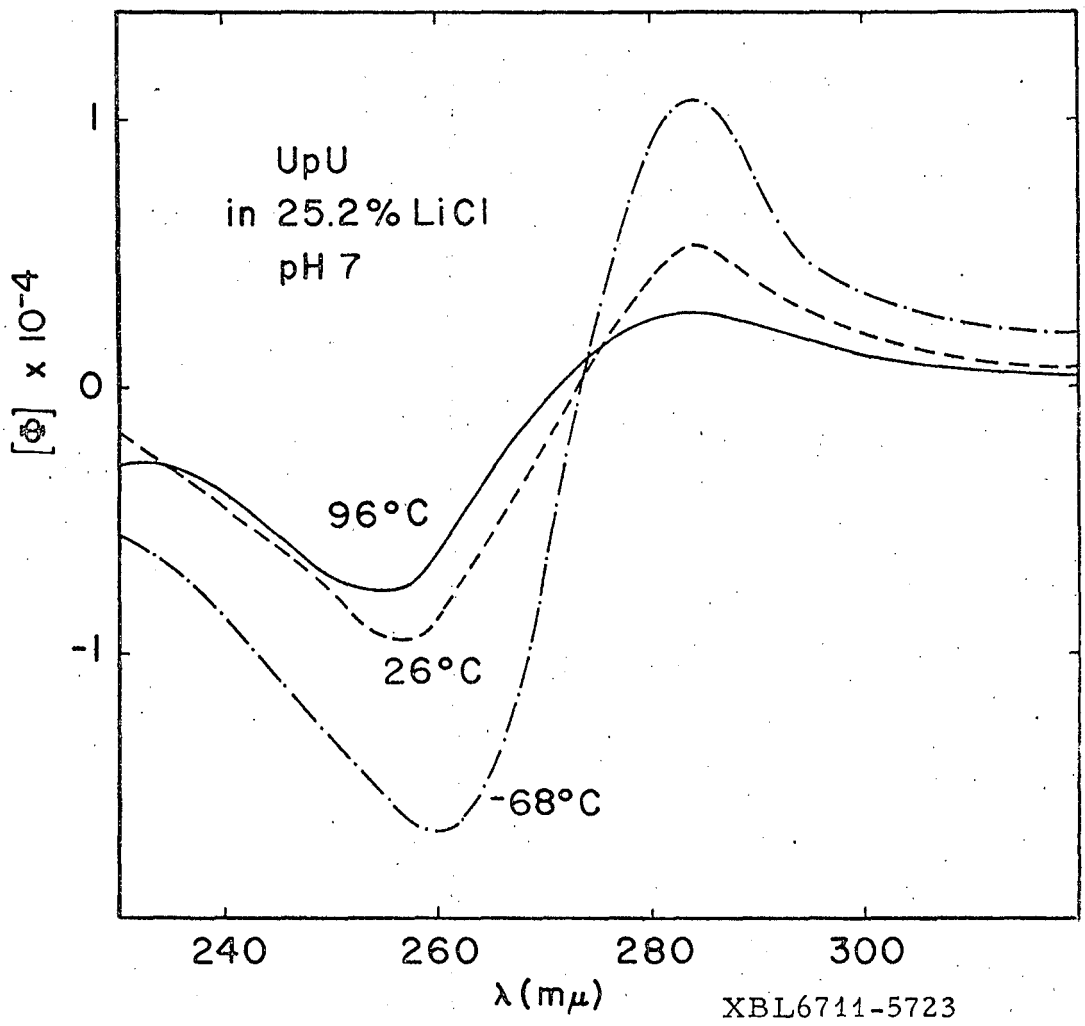



Figure 25 ORD of UpU in 25.2% LiCl solution at -68°, 26° and 96°C.

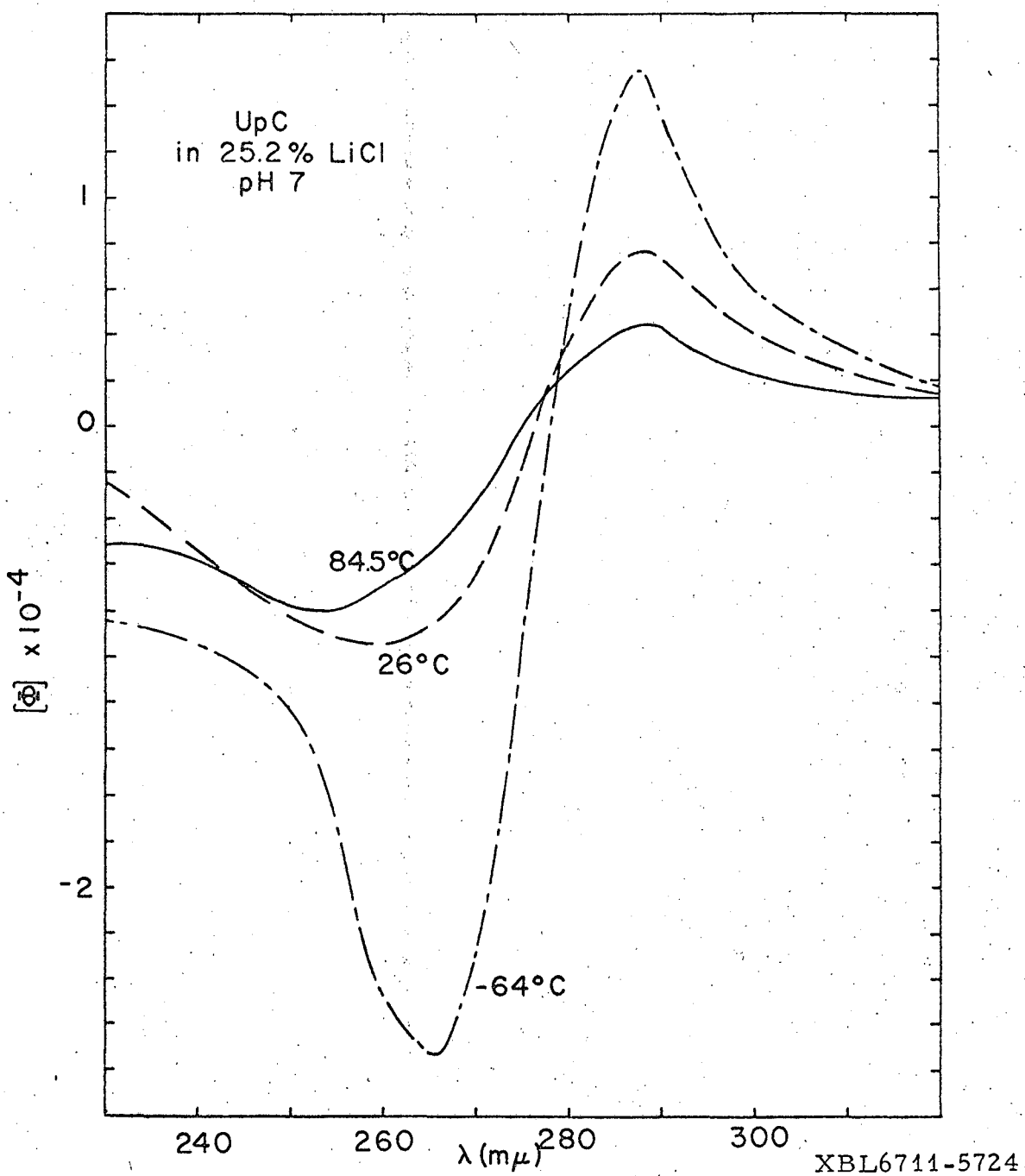



Figure 26 ORD of UpC in 25.2% LiCl solution at -64°, 26° and 84.5°C.

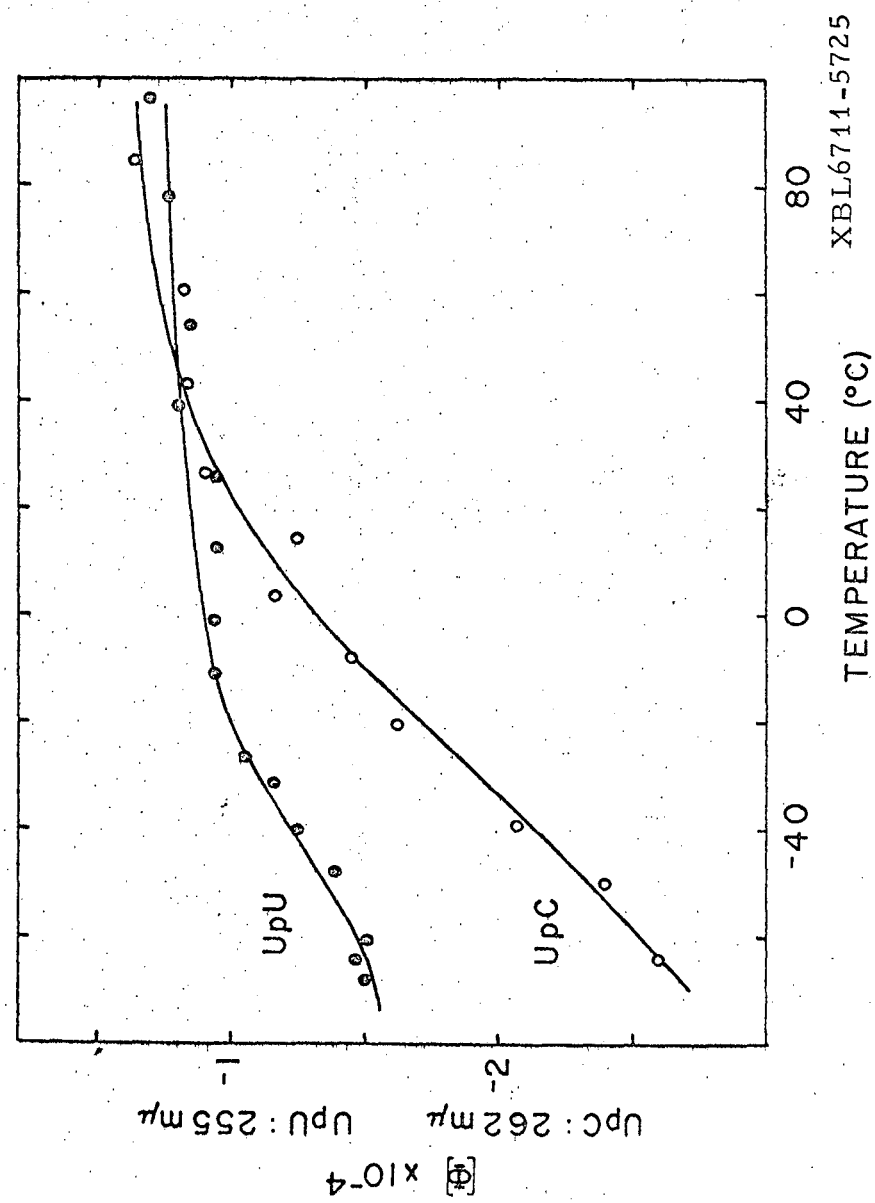



Figure 27 $[\Phi]_{265}$ of UpU and UpC in 25.2% LiCl solution as a function of temperature.

pretation of the data for UpU is open to question because UpU has considerable hypochromicity in 25.2% LiCl. $\epsilon_{max}$ of U decreases from 10000 in dilute salt buffers to 9400 in 25.2% LiCl while $\epsilon_{max}$ for UpU decreases from 9800 to 7300 (at about $10^{-4}$M base/liter). Also, the absorption of UpU (see Table 1) does not obey Beer's Law at room temperature. This indicates aggregation. However, the ORD of UpU in 25.2% LiCl is very much like the ORD in dilute salt solutions. There is the possibility that the increase in optical rotation which occurs only well below 0°C is caused by aggregation rather than an increase of the intramolecular base-base interaction. On the other hand UpC does not display any appreciable additional hypochromism when dissolved in 25.2% LiCl (even at low temperatures). The evidence in the latter case strongly indicates that aggregation is not taking place and that the increase in the magnitude of the ORD at low temperatures is caused by intramolecular interactions. The temperature dependences of the ORD of these dinucleoside phosphates are very different from ApA or CpC. The midpoint in the transitions of the ORD of ApA and CpC are approximately +5°C (Figures 18,24). In the case of UpC and UpU this temperature is -30°C or lower (Figure 27). Expressing the difference in another way, at 25°C the interaction between the two bases responsible for the change in the ORD is about 1/3 of the low temperature limit for ApA and CpC while in the case of UpU and UpC it is less than 1/10. Our experiments have not enabled us to say whether UpU and UpC at room temperature are like ApA

at 90°C, (where the bases remain parallel to each other although the structure is quite flexible) or whether there is an appreciable fraction of the molecules "unstacked" with the two bases constrained only by the covalent bonds connecting them.

In either case we expect polyuridylic acid or the alternating polymer of uridylic acid and cytidylic acid to be quite different from polyadenylic acid or polycytidylic acid. The former ones would be much more flexible and much less rigid than the latter ones. This agrees with previous hydrodynamic measurements[162]. This suggests that hairpin turns (in molecules such as transfer RNAs) would be more likely to occur in regions containing sequences such as ...pUpUp... or ...pUpCp...

COMPARISON WITH THE LITERATURE

In reaching a reasonably refined model of the conformations of dinucleoside phosphates in solution almost all of the data used were ours. There is, as we have already noted, a lot of work reported in the literature which also treats this problem. On the basis of their work, other investigators have reached conclusions of their own. Comparisons of our results with theirs will reflect the reliability of both the experimental data and the methods used to analyze them.

The first group of experiments measured the optical properties (ORD, CD and ultraviolet absorption) of dinucleoside phosphates and related oligomers. In one series of experiments the ORD and ultraviolet absorption of dinucleoside phosphates were measured in neutral, acidic and alkaline

solutions at room temperature[208,210,211]. The conclusions reached in that study were that dinucleoside phosphates with both bases electrically charged or with at least one base being uracil tended to be considerably less "stacked" than the other dinucleoside phosphates. Measurements of the ORD of dinucleoside phosphates in 25.2% LiCl agree with this. ApA and CpC may be considered to be typical of dinucleoside phosphates which "stack" readily while UpU and UpC may be considered to be representative of dinucleoside phosphates which don't have strong base-base interactions. However, closer comparisons do not agree as well. Table 5 compares results from that work and this. (The first three columns are taken from reference 44, Table III-5.) Six criteria for measuring the amount of "stacking" in dinucleoside phosphates are listed. They are hypochromism, hypochromicity at the maximum of the absorption (near 260 mμ), the maximum difference in optical rotation between a dinucleoside phosphate and its constituent nucleotides, the thermodynamic parameters for the "unstacking" process (determined using the two-state model), $\Delta H^{o}$ and $\Delta F^{o}$ (25°C), and the torsional force constant, κ, determined using the oscillating dimer model. The dinucleoside phosphates at the top of each column are those which seem most "stacked" according to that particular criterion. Large values for hypochromism, hypochromicity and optical rotation are considered indicative of "stacking". Large positive values of $\Delta H^{o}$ ("unstacking"), $\Delta F^{o}$ ("unstacking") and κ are also considered indicative of "stacking".

TABLE 5

DINUCLEOSIDE PHOSPHATES IN ORDER OF INCREASING "STACKING"

| Hypochromism | Hypochromicity | ORD | $\Delta H^{o}$ | $\Delta F^{o}$ | $\kappa$ |
|---|---|---|---|---|---|
| GpG | CpG | ApA | ApU | ApU | ApU |
| GpC | GpC | CpC | UpU | UpU | CpC |
| ApC | ApA | ApC | GpC | GpC | ApC |
| ApA | CpA | CpU | CpU | CpU | GpC |
| GpA | ApC | CpA | CpA | CpA | UpC |
| UpG | GpA | ApG | CpC | GpU | UpA |
| CpA | CpC | UpG | GpU | CpC | CpU |
| CpU | GpG | ApU | UpC | UpG | UpU |
| CpC | CpU | GpA | ApC | ApC | ApG |
| ApG | ApG | GpU | UpG | UpC | CpG |
| CpG | UpG | CpG | GpA | UpA | CpA |
| ApU | ApU | CpG | ApA | GpA | ApA |
| UpA | UpA | GpG | UpA | ApA | GpA |
| UpC | UpC | UpU | ApG | ApG | GpU |
| GpU | GpU | GpC | CpG | CpG | UpG |
| UpU | UpU | UpA | | | |

Correlation between columns is not good. This reflects the fact that there is no single measurement which quantitatively describes the extent of base-base interactions. A single ORD measurement is probably not as dependable as ultraviolet absorption or NMR measurements because of its strong angular dependence. The correlation of the order of dinucleoside phosphates according to hypochromism, $\Delta H^{o}$ and $\kappa$ is almost random. This indicates that good estimates of the properties of fully "stacked" dinucleoside phosphates are needed in order to make reliable estimates of the amount of base-base interactions at room temperature, for example. Measurements of more physical properties (e.g. ORD, ultraviolet absorption, and NMR) of more dinucleoside phosphates in concentrated salt solutions down to low temperatures will be a great help to us in estimating the conformations of dinucleoside phosphates at higher temperatures.

The temperature dependences of the CD of a number of dinucleoside phosphates have been measured[14-16]. The two-state model was used to analyze those data. A comparison between those values and ours is shown in Table 6.

The average deviations for $\Delta H^{o}$ and $\Delta S^{o}$ as determined in the two cases are reasonably small, 1 kcal/mole and 3.7 cal/deg. mole respectively. The relative differences in $\Delta F^{o}$ are considerably larger, making it difficult to compare the degree of "stacking" measured in these two ways. Considering the difficulty of the experiments and the independent estimations of high and low temperature

TABLE 6

COMPARISON OF $\Delta H^{o}$ AND $\Delta S^{o}$ FOR "UNSTACKING" OF DINUCLEOSIDE PHOSPHATES

| Dinucleoside Phosphate | ORD | | CD[14-16] | |
|---|---|---|---|---|
| | $\Delta H^{o}$ (kcal/mole) | $\Delta S^{o}$ (cal/deg. mole) | $\Delta H^{o}$ (kcal/mole) | $\Delta S^{o}$ (cal/deg. mole) |
| ApA | 5.3 | 20 | 8.0 | 28 |
| ApC | 6.2 | 22 | 6.1 | 21 |
| ApU | 8.4 | 32 | 6.7 | 24 |
| CpA | 7.3 | 27 | 7.0 | 24 |
| CpU | 7.8 | 28 | 6.8 | 24 |
| CpC | 6.9 | 25 | 7.5 | 25 |
| GpA | 5.6 | 20 | 6.1 | 22 |

limits of the CD and ORD the differences could easily come from experimental errors and parameters chosen for the analysis, rather than an inherent difference in the measurements. It is unfortunate that the largest difference occurs in the case of ApA, the first dinucleoside phosphate to be analyzed. This implies that the difference in $\Delta H^{o}$ as determined from CD and ORD measurements is a result of difficulties in the experimental measurements and analysis, rather than something inherent in the "unstacking" process. This fact makes it seem likely that $\Delta H^{o}$ of ApA as determined from CD data should agree with those determined from ORD data, rather than those determined from ultraviolet absorption measurements. (see Table 4) It should be noted that if the shape of the ORD (or CD) caused by the interactions between the bases changes with changing temperature then we would not expect the answers to be the same. Although a difference between the temperature dependence of the magnitude of the 260 mμ trough and the 213 mμ trough of the ORD of ApA has been noted[187] no changes in shape in the region of the spectra which are analyzed (Cotton effects near 260 mμ) have been found.

As noted before (see Table 4) there is an appreciable difference between $\Delta H^{o}$ ("unstacking") for ApA as measured by ORD and ultraviolet absorption measurements (as well as calorimetric measurements which use ultraviolet absorption data in their analyses. Since these differences may very well be real (ORD and ultraviolet absorption have different geometric dependences.) comparisons of the conclusions

reached by the two methods reflect not only experimental error, but differences inherent in the problem.

Osmotic coefficient measurements and NMR spectroscopy of nucleosides and bases have enabled investigators to compare tendencies to "stack" in these systems[21,33,35,169,176,194,196]. The greatest amount of "stacking" is found where adenine is the base, the least where uracil is the base. This agrees with our work. Measurements of the temperature dependence of the ORD of dinucleoside phosphates in 25.2% LiCl solutions show ApA to be far more "stacked" than UpU or UpC. However these ORD measurements also show CpC to be as "stacked" as ApA. NMR and osmotic coefficient measurements indicate that cytosine and cytidine "stack" considerably less than do adenine and adenosine. We must remember that the "stacked" conformation in dinucleoside phosphates may be quite different from the "stacked" conformation in the case of bases or nucleosides. The phosphodiester linkage makes it impossible to orient the bases of a dinucleoside phosphate in the same way as it is proposed that purine ribosides "stack"[21]. NMR studies of dinucleoside phosphates, on the other hand, would serve as a very useful comparison to optical measurements.

The last comparisons are with the calorimetric and statistical mechanical estimates of the thermodynamic parameters for the "unstacking" process[40,66,179]. Values of $\Delta H^o$ do range from 2 to 7 kcal/mole which are compatible with our estimates. However further comparisons are limited by the

same difficulties that were encountered previously in comparing results from bases and nucleosides with those from dinucleoside phosphates, the problem of not knowing whether the process being measured is really the same in both cases.

ENERGETIC CONSIDERATIONS

We have discussed extensively the conformations of dinucleoside phosphates and the variation of these conformations with changes in the sequences of the dinucleoside phosphates or changes in their environments. From this information we can say something about the origins of the forces stabilizing these structures. In a thermodynamic problem such as this, the word "forces" takes on a special meaning. Not only do we have to consider the energies (enthalpies) of conformations; but we must also consider their entropies. The total free energy, $F = H - TS$, determines the stability of any given conformation. The origins of the enthalpic part of the free energy are easy to see. They include electrostatic interactions, hydrogen bonds and steric repulsions. The origins of the entropic contributions to the free energy are not understood as well. They include solute entropy terms such as free or hindered movement about single bonds and probably most important the solvent entropy terms which include ordering and disordering of the solvent. Because these enthalpic and entropic effects are sometimes difficult to separate from one another experimentally it is easier to discuss experimental results and theoretical calculations if we group these interactions by the parts of the dinucleo-

side phosphates which are participating. There are interactions involving the bases, the sugars and the phosphate group. In addition, interactions with the solvent must be considered.

The interaction most widely discussed is the one between bases. The fact that purine and pyrimidine bases aggregate in aqueous solutions indicates that there are strong interactions between bases. "Stacking" interactions of purine and pyrimidine bases in solution are not unique though. The aggregation of water-soluble aromatic molecules such as methylene blue and acridines has been studied for a very long time[11,78,103,109,110,131,137,138,155,174,221,222]. The aggregation of these molecules is driven by a favorable change in enthalpy as is the case for bases, nucleosides and dinucleoside phosphates, with $\Delta H^{o}$ ("stacking") being about -7 kcal/mole. In looking for theorigin of "stacking" we should concentrate on properties common to all these molecules.

The tendency to aggregate for the series of purine nucleosides studied correlates well with the polarizabilities of the bases and not the dipole moments[21]. (See Table 7.) This points to dispersion interactions as the major stabilizing interactions (as opposed to dipole-dipole interactions, for example) because dispersion interactions depend most strongly on the magnitude of the polarizability. The London dispersion interaction is approximately:

TABLE 7

| COMPOUND | μ(Debye) | α(Å$^3$) |
|---|---|---|
| Caffeine | 3.4 | 19.3 |
| 6-Dimethylaminopurine | 3.3 | 17.6 |
| Guanine | 6.8 | 14.4 |
| 6-Methylpurine | 3.8 | 14.3 |
| Adenine | 3.2 | 13.9 |
| Hypoxanthine | 5.2 | 13 |
| 5-Bromouracil | 4.5 | 12.9 |
| Purine | 4.2 | 12.5 |
| Thymine | 3.6 | 12 |
| Cytosine | 7.2 | 11 |
| Uracil | 3.9 | 10.2 |

These are theoretical values, calculated in reference 151.

$$E_{AB} \cong \frac{3}{2} \frac{I_a I_b}{I_a + I_b} \frac{\alpha_a \alpha_b}{r_{ab}^{\ 6}}$$

where $I_a$ and $I_b$ are the ionization potentials of groups a and b, $\alpha_a$ and $\alpha_b$ are the polarizabilities of a and b, and $r_{ab}$ is the distance between a and b. This interaction comes about because of fluctuations of the electron distribution in one group which creates temporary dipole which in turn polarizes the charge distribution of the second group, creating a net attraction[120]. There are also dispersion forces caused by fluctuations which create quadrupoles and even higher electric moments. These interactions fall off

even more rapidly (e.g. $r^{-8}$ for the quadrupole term). The formula for the interaction is only approximate in that it is derived for molecules A and B far apart. In the case of "stacking" the distance between bases is essentially the minimum contact distance so that serious errors may be introduced by this approximation. Calculations of the London dispersion interaction for bases in DNA geometry have been made[36,48,152,153]. The calculated values of this attraction are in the same range as $\Delta H^{o}$ ("stacking"). This supports the hypothesis that dispersion forces may well dominate the base-base attraction.

It must be remembered that the correlation of polarizability with "stacking" ability was made in the case of purine nucleosides[21], where there were no covalent ribose-phosphate connections between bases. This covalent linking of the bases in dinucleoside phosphates severely limits the range of conformations available to the bases. It may very well be that the "stacked" conformation of bases found in nucleosides may be forbidden in the case of dinucleoside phosphates for steric reasons. (This is the case for a model of "stacked" nucleosides proposed[21].) It would be good if such a correlation of "stacking" with polarizability could be demonstrated or disproven in the case of dinucleoside phosphates. In order to do this low temperature measurements similar to those already made for ApA would have to be done for a very large number of dinucleoside phosphates. We do know that ApA (consisting of two bases with high polarizabilities)

is one of the most highly "stacked" dinucleoside phosphates while UpC and UpU (consisting of bases with low polarizabilities) are among the least "stacked". However CpC seems to be as "stacked" as ApA. Cytosine has a considerably lower polarizability than does adenine, but has a high dipole moment. This indicates that dispersion interactions do not completely dominate other forces in the case of dinucleoside phosphates, that other electrostatic forces may be important as well. The large dipole moment of cytosine may cause large dipole-dipole or dipole-polarizability interactions, for example.

Additional evidence for the importance of dispersion forces comes from the effects of substitution at the $C_5$ position of pyrimidines. Substitution of a bulkier and more polarizable group for $H_5$ stabilizes "stacked" structures. This applies for substitutions of I and Br as well as methyl groups in the $C_5$ position of both cytosine and uracil. Both single- and multiple-strand polymeric structures seem to become more ordered, as indicated by an increases in hypochromicity and ORD and an increases in the $T_m$ as compared to polynucleotides containing the parent bases with an $H_5$ [123,185,186]. In addition, NMR and osmotic coefficient measurements indicate that bases, nucleosides and dinucleotides with these bulkier and more polarizable groups attached to $C_5$ of a pyrimidine "stack" more readily than do cytosine and uracil (or cytidine and uridine, etc.[21,33]).

There are two reasonable explanations for this. The $C_5$ substitution group (i.e. methyl, I, etc.) is directly under

the base of its 5' neighboring nucleotide. This polarizable group can interact with the adjacent base via dispersion forces to stabilize a "stacked" structure. Another possible explanation is that this $C_5$ substitution group may add to solvent stabilization by either enthalpic (surface tension) or entropic (hydrophobic) mechanisms. These types of solvent interactions will be discussed later.

The first choice (dispersion interactions) seems more reasonable. A Br attached to the pyrimidine $C_5$ stabilizes structures more than does a methyl group. Both groups are very close to the same size[12]. However Br is less hydrophobic and more polarizable. This seems to support the contention that dispersion interactions (which depend strongly on the polarizability) are important here and in the cases of interactions between other bases.

The same calculations which suggest large contributions to the stability of "stacked" bases in nucleic acid structures also show large contributions from other electrostatic interactions[36,48,152,153]. These include charge-charge interactions (or if the charge distribution of the electrically neutral base is expressed as a point dipole, dipole-dipole interactions) and charge-polarizability (or dipole-polarizability) interactions. These must be considered as well in evaluating interactions between bases.

These non-dispersive electrostatic forces also involve the other groups of the dinucleoside phosphate. The sugars and phosphates do not have large planar areas with high

polarizability (as do the bases) which can align themselves parallel with large dispersion forces stabilizing this conformation. The sugars and phosphate do contain groups with large charges or dipole moments (e.g. hydroxyl groups and the phosphate group). These groups can interact strongly with one another or with the base. Two specific (and mutually exclusive) interactions of this type have been proposed. They are both invoked to explain the difference in properties of ribo- and deoxyribopolynucleotides [31,32,147,197,206]. One hypothesis proposes that the 2' OH of the ribose is hydrogen bonded to either the $N_3$ of purines or the $C_2$ carbonyl oxygen of pyrimidines[21,197]. This hydrogen bond is claimed to help make ordered helical structures of ribopolynucleosides more stable than the similar structures of their deoxyribo analogues[197]. Experimental evidence for such a hydrogen bond is discussed at length in reference 197. It includes infrared, NMR and protonation studies on polynucleotides and model compounds. The evidence is by no means straight forward. We feel that such a hydrogen bond is unlikely for several reasons. Although this proposed hydrogen bond can be made using Courtauld space-filling atomic models the bond is quite bent. This would greatly reduce the energy gained by forming it (as opposed to linear hydrogen bonds)[139]. The 2' OH can hydrogen bond with a number of hydrogen acceptors, including water, the 3' OH, the 5' OH or phosphate oxygens. There doesn't seem to be any reason to form a hydrogen bond preferentially with the purine or pyrimidine base rather than any

of the other choices. In addition, X-ray diffraction studies of both single crystals[23,57,58,83,99,173,181,182] and fibers[4] show no evidence of such a bond. Finally, if such a hydrogen bond were made there is no obvious reason why it should stabilize ordered helical structures. It would make the nucleotide itself more rigid by limiting motion about the glycosidic bond. However, it is hard to see how this would influence internucleotide interactions, especially those between bases.

Another suggestion is that the 2' OH is hydrogen bonded to a phosphate oxygen[14,15]. The choice of this explanation of the differences between 3'-5' linked oligoribonucleotides and 2'-5' linked oligoribonucleotides or oligodeoxyribonucleotides is really one of expedience. There is no evidence cited which supports this contention directly. The authors consider other explanations even less adequate. Other explanations of the differences between the properties of ribo- and deoxyribopolynucleotides will be discussed in the second part of this thesis.

The problem of interactions of various parts of the ribose-phosphate backbone are treated extensively in the second part of this thesis. Even though the choice of formalism may vary in approaching this problem, the treatment remains essentially the same. There are a large number of interactions between charged and/or polarizable groups or atoms. They must be calculated for all of the many possible conformations. Depending on the cleverness of the investigators

trying to solve this problem, it may or may not turn out to be a "brute force" type of computer calculation. The results of just such calculations will be discussed later.

The interactions mentioned so far contribute to $\Delta H^o$ of "stacking" and not $\Delta S^o$. A large part of $\Delta S^o$ for this transition comes from the increase in flexibility (or randomness) as a result of "unstacking" a dinucleoside phosphate. There are seven single bonds about which reasonably free rotation can take place when the bases are not held close to one another, but which are greatly hindered when bases are "stacked". These bonds are the two glycosidic bonds, the $C_3'$-$O_3'$ bond, the $O_3'$-P bond, the P-$O_5'$ bond, the $O_5'$-$C_5'$ bond and the $C_5'$-$C_4'$ bond. Neglecting the potential energy for rotation about these bonds, the entropy gained by the increase of configurations available is

$$\Delta S = R \ln \frac{\omega_u}{\omega_s}$$

where R is the gas constant and $\frac{\omega_u}{\omega_s}$ is the ratio of angular configurations allowed (about a given single bond) in the "unstacked" and "stacked" dinucleoside phosphate. If $\frac{\omega_u}{\omega_s}$ is 3 then there is a contribution of approximately 2.2 cal./degree mole for each of these bonds. Multiplying this value by the number of bonds, seven, we see that a contribution of 15 cal./degree mole is not unreasonable. Since the values of $\Delta S^o$ ("unstacking") are about 20 to 30 cal./degree mole, it is quite possible that this internal configurational entropy is the dominant term contributing

to the entropy of "unstacking".

Up until this point we have judiciously avoided discussing effects of the solvent on the structure of dinucleoside phosphates. Water is unique in its ability to stabilize "stacked", helical polynucleotide structures[43,72,175]. Although the bases found in nucleic acids are aromatic and to a large part hydrophobic, hydrophobic bonds (which are responsible for much of the secondary and tertiary structure of proteins) are entropic in nature[95,167]. The driving force for this type of bonding comes from the disordering of the solvent when hydrophobic groups are brought together. Increasing the temperature increases the importance of entropy changes on the equilibria among conformations and the hydrophobic bond becomes stronger. This is not the case with single-strand oligonucleotides and polynucleotides. The hydrophobic groups (i.e. the bases) interact less strongly and become more exposed to the solvent with increasing temperature.

Sinanoglu and Abdulnur[175] have developed a theory of liquid structure around solvophobic solutes which predicts an increase in enthalpy upon bringing purine and pyrimidine bases together. The crux of their argument is that a large base-solvent interface has an unfavorable enthalpy associated with it, caused by the surface tension between the two "phases". "Stacking" of bases reduces the interfacial area and is therefore associated with a favorable change in enthalpy. Some of our data indicate that water structure around the

bases may be important in stabilizing certain conformations, but with a slightly different mechanism.

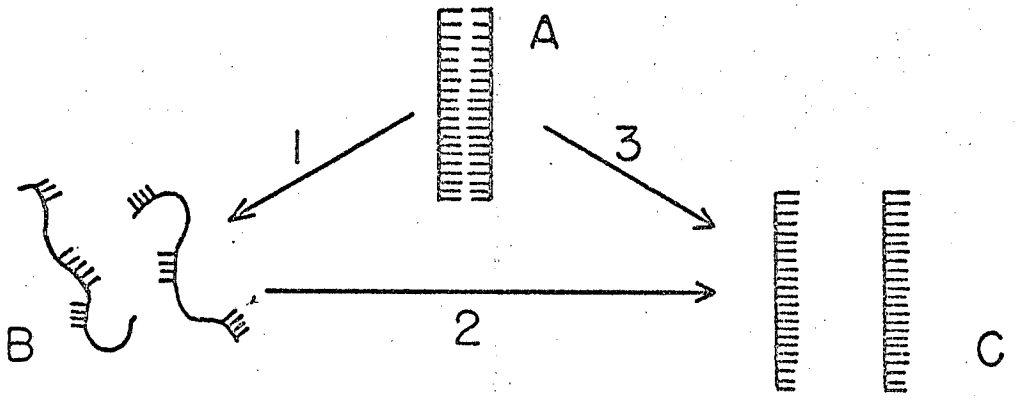
There is one interesting correlation of the $\Delta H^{o}$ of "unstacking" and the apparent torsional force constant for different dinucleoside phosphates (Table 8). Dinucleoside phosphates seem to fall into two groups, those with a purine nucleoside attached by a 5' ester and those with a pyrimidine nucleoside attached by a 5' ester. This has some implications as to the forces responsible for stabilizing their conformations. Apparently the key factor is not simply the overlapping area of the bases, in which case one would expect the purine-purine dinucleoside phosphates to have the largest values. Pyrimidine bases, nucleoside and nucleotides are considerably more soluble than their purine counterparts. This is probably caused by strong hydration. The carbonyl and amino groups probably interact with water most strongly. If the low temperature form of the dinucleoside phosphate starts a right handed helix with both bases in the anti orientation about the glycosidic bond[50] (these are the orientations most commonly found in crystal and fiber structures) then the carbonyl and amino groups of a 3' pyrimidine would be over the other base, while the carbonyl and amino group of the 5' pyrimidine would be exposed to the solvent. Strong interaction with the solvent is much more likely in the latter case. This interaction might serve as a nucleus of some solvent structure which would lead to a large $\Delta H^{o}$ ("unstacking") and large force constant for the oscillating dimer model.

TABLE 8

| | $\Delta H^o$ "unstacking" (in kcal/mole) | | $\kappa$ (in cal/mole radian$^2$) | |
|---|---|---|---|---|
| | Average | Range | Average | Range |
| Purine-Purine | 5.2 | 5.6-4.8 | 200 | 230-170 |
| Pyrimidine-Purine | 5.8 | 7.3-4.8 | 200 | 250-130 |
| Purine-Pyrimidine | 7.3 | 8.4-6.2 | 260 | 340-140 |
| Pyrimidine-Pyrimidine | 7.2 | 7.8-6.2 | 260 | 300-230 |

It is interesting to contrast this with the theory of Sinanoglu and Abdulnur[175]. Although they predict the solvent plays the key role in stabilizing the "stacked" conformation the dominant features influencing this stability is the size and shape of the bases, not their ability to interact strongly with the solvent at one or a few specific sites.

There is one important case where entropic (hydrophobic) forces play a very important role in stabilizing ordered polynucleotide structures. Let us consider the following cycle of equilibria:




A is a double-strand complementary base-paired helix; B are the two single-strand polymers at $T_m$ where they are partially "stacked"; C are the single-strand completely "stacked" polynucleotides. Reactions 1, 2, and 3 form a cycle so that

$$\Delta H_1^o + \Delta H_2^o = \Delta H_3^o$$

and

$$\Delta S_1^o + \Delta S_2^o = \Delta S_3^o$$

We can estimate $\Delta H_1^o$, $\Delta H_2^o$, $\Delta S_1^o$, and $\Delta S_2^o$. Calorimetric measurements of $\Delta H^o$ are about 4 kcal/mole base (8 kcal/mole base pair)[1,24,149,163]. The $T_m$ of helix-coil transitions

of polynucleotiees ranges a bit but 80°C (353°K) can be considered a representative value[126]. At this temperature $\Delta F_1^o = 0$ so that $\Delta S_1^o = \Delta H_1^o/T_m = 11.3$ cal./degree mole base. $\Delta H_2^o$ is determined from calorimetric and optical measurements of the "stacking" process in single-strand polynucleotides. If we take as representative values -8.0 kcal/mole base for $\Delta H^o$ ("stacking") and -27 cal./degree mole base for $\Delta S^o$ ("stacking")[3,14-16,56,106,147,203,206], and consider the single-strand polynucleotides at $T_m$ to be 20% "stacked"[3,14-16,56,106,147,203,206] then $\Delta H_2^o = -6.4$ kcal/mole base and $\Delta S_2^o =$ $= 21.6$ cal./degree mole base. Adding reactions 1 and 2 together we conclude that $\Delta H_3^o = -2.4$ kcal/mole base and $\Delta S_3^o = -10.3$ cal./degree mole base! That means that the transition from the single-strand "stacked" form to double-strand form in polynucleotides is driven by a favorable change in entropy, not enthalpy. Apparently the decrease in exposure of the hydrophobic bases to the solvent in the process of base-pairing forms strong hydrophobic bonds. Hydrogen bonds between complementary bases are formed at the expense of those between the bases and water. It is difficult to say which hydrogen bonds are stronger. Because the bond energy of a hydrogen bond is small the change in enthalpy caused by the exchange of hydrogen bonds is also small, probably $0 \pm 1$ kcal/mole hydrogen bond[139].

It would be interesting to look for such a transition. Assuming that our estimates are correct then the transition temperature for reaction 3 (the temperature at which

$\Delta F_3^o = 0$) is about -40°C (233°K). Our measurements on dinucleoside phosphates show that at -40°C single-strand polynucleotides would not be completely "stacked". This would cause the transition temperature to be lowered. However, -40°C is well above the freezing point of 25.2% LiCl. Such a transition may very well be experimentally observable. Of course, our knowledge of the thermodynamic parameters used in this analysis is meager. It would not be at all surprising if some of these values were in error by as much as 20%, which means that the transition temperature might be too low to observe.

We have seen that there is no single interaction which dominates other forces in the case of dinucleoside phosphates and single-strand polynucleotides. There are important contributions from dispersion interactions, other electrostatic interactions and solvent interactions, both enthalpic and entropic. This makes it difficult to predict the different conformations of dinucleoside phosphates as well as the effects of sequence on these stabilities. It does not seem possible to pick one or a small number of properties of the bases, then describe the structure of dinucleoside phosphates in terms of them. In order to calculate a detailed multi-dimensional energy surface (then calculate the temperature dependence of physical properties so that we can test it) we must take into account all of these interactions. The second section of this thesis is devoted to just such a calculation.

CONCLUSION

A number of conclusions can be drawn from this work. First of all, the unstacking process is not a two state process. This means that the thermodynamic parameters obtained from the analysis of the temperature dependence of the optical properties may or may not be meaningful. They ceratinly should be considered with a great deal of caution.

In aqueous solutions the bases of a dinucleoside phosphate tend to be parallel, above one another with no solvent between them, even at high temperatures. The bases probably oscillate back and forth, with the magnitude of the oscillations increasing with increasing temperature.

The degree of interaction between the bases of a dinucleoside phosphate may vary greatly from one compound to another. The interaction between the bases in ApA or CpC is considerably greater than in UpU or UpC. The latter may have the same general conformation, but with larger oscillations of the bases. At room temperature and above these uracil containing dimers may have slight base-base interaction.

Since the helix to coil transition in single-strand polynucleotides seems to be essentially non-cooperative[3,16,106,147,203,206] we expect the base-base interactions in single-strand polynucleotides to be similar to those in dinucleoside phosphates. Therefore, a single-strand polynucleotide at the mid-point of its helix-coil transition

would tend to have all of its bases half-stacked rather than half of its bases fully stacked and half fully unstacked. This means that there might be a considerable amount of short range disorder while still retaining a fair amount of long range order. In aqueous solutions single-strand polynucleotides such as polyadenylic acid are probably more like wormlike coils rather than a series of stiff rods connected by universal joints.

In a polynucleotide such as polyuridylic acid or one containing alternating uridylic and cytidylic acid residues there should be important differences. Our experiments have not enabled us to say whether UpU at room temperature is like ApA at 90°C, (where the bases remain parallel to each other although the structure is still quite flexible) or whether there are an appreciable fraction of the molecules unstacked with the two bases constrained only by the covalent bonds connecting them. In either case, we would expect to find hairpin turns occuring more frequently in regions containing sequences such as ...pUpUp... or ...pUpCp...

In addition, single-strand RNAs are probably dynamic structures on a very short time scale, much shorter than implied by the tritium exchange studies on DNA and transfer RNA[55,148]. Although single-strand regions of RNA are stiff, they will also be in constant, rapid motion.

# CALCULATIONS OF STABILITIES OF NUCLEIC ACID STRUCTURES

## INTRODUCTION

The interpretation of the experimental measurements of temperature-dependent properties of dinucleoside phosphates has left the question of dinucleoside phosphate structure unresolved. The experimental results which we have already discussed indicate that there is an ordered low temperature form (or forms) of dinucleoside phosphates in which bases interact with the adjacent nucleotide. At high temperature this interaction decreases, although in the case of ApA (and probably other dinucleoside phosphates as well) there is still appreciable base-base interaction at 90°C. In addition, we know that neither a two-state model or the harmonic oscillator model is compatible with the measured temperature-dependent properties of dinucleoside phosphates More realistic (and therefore more complex) geometries and energy surfaces, are needed to explain the experimental results. In this part of the thesis we will describe calculations of the energies of conformations of nucleic acid materials which give us a starting point for developing new models.

If we are successful in accomplishing this first objective then a second objective becomes accessible. If we can show that a computational technique gives reliable energies of conformations of nucleic acid material, why would it not be possible to calculate *a priori* nucleic acid structures? We could address ourselves to problems such as: How hard is

it to make a hairpin turn? or What are the stabilities of the "clover leaf" structures proposed for transfer RNAs? or What is the structure of the RNA double helix? In spite of the fact that we have no guarantee that this technique will be fruitful, it should enable us to suggest possible structures (which can then be looked for experimentally) and to comment on the origins of the differences in structure between analogous compounds (e.g. ribo- and deoxypolynucleotides).

In order to have confidence in the results of these calculations their reliability must be demonstrated. The easiest comparisons with experimental data which can be made are those with structures determined by X-ray crystallography. These comparisons are not the best ones to make, as we are interested in the structure of isolated molecules in aqueous solution, not in the solid state. In spite of the fact that the crystal structures of proteins (e.g. myoglobin) are very similar to the structures of these molecules in solution there are extensive interactions (such as intermolecular hydrogen bonds and crystal packing forces) which may appreciably perturb the solution conformations of small nucleic acid fragments. These may lead to differences between the conformation of a molecule in the crystal and the conformation of minimum potential energy which we calculate for the isolated molecule. However, the molecular conformation found in the crystal should have an energy which is not much higher than the energy of the most stable conformation of the isolated molecule.

A more reasonable test would be a comparison of the conformation-dependent properties (i.e. ORD, NMR, etc.) of the molecule in solution with those calculated using the geometry of the lowest energy conformation predicted. Such calculations of ORD are presently being undertaken[84,90]. Another approach is to compare the properties of molecules under consideration with model compounds of known conformation. For example, comparisons of the ORD of nucleosides and their analogues have led to the determination of the orientation of the glycosidic bonds for some nucleosides[53,54,98,128,201,219]. It must be remembered that these comparisons with experimental results will become more difficult as the compounds under investigation become larger. In order to be able to calculate structures of larger molecules with any degree of confidence we must be very sure that our results for the smaller molecules are correct.

Implicit in this discussion is the assumption that the molecules of interest exist in the conformations of lowest energy in their biologically active forms. It is quite possible that as a polymer starts "folding" it becomes trapped in some local free energy minimum, never having a chance to reach the true minimum. Some studies of proteins indicate that this may not be the general case. It has been shown that the two different chains of hemoglobin are synthesized separately and diffuse together in order to form the complete molecule[5]. During the time when the chains are separated they do not become trapped in an energy minimum

which prevents them from aggregating properly. In other studies ribonuclease, aldolase and alkaline phosphatase have been completely denatured (using urea and reducing agents to break disulfide bonds) and then renatured in vitro (dialyzing out the urea and reoxidizing the sulfhydryl groups to form disulfide bonds)[2,108,178a]. Obviously the polypeptide chains fold properly to form the native enzymes (presumably the energy minimum) and do not become trapped in denatured conformations which are local energy minima. These studies suggest that all the information necessary to determine the structure of these proteins is contained in the amino acid sequence. However, these are relatively small proteins. More complicated biopolymers may not be able to undergo this type of reversible denaturation.

In the case of molecules no larger than dinucleoside phosphates (and considerably smaller than ribonuclease or alkaline phosphatase) it is likely that the energy barriers separating conformations will be low and that the exchange between conformations will be rapid. We will demonstrate this explicitly in the case of CpC.

Therefore we will try to compute the conformations (or distribution of conformations) of nucleic acid fragments by calculating their energies as a function of varying conformations as these conformations are determined by thermodynamics rather than kinetics. Many similar calculations have been attempted, from which we can learn a great deal.

There have been two general approaches used in attempts to find the most stable conformations of polypeptides,

polynucleotides and synthetic organic high polymers. We will call these two approaches microscopic and macroscopic. Most of the work done up to this time has been based on the microscopic approach[10,17-20,45,64,65,115,129,135,156-158,166,168,171,202]. In this approach the polymer is considered to consist of a series of rigid moieties, connected by single covalent bonds about which reasonably unhindered rotation can take place. (In the case of nucleic acids these would be the bases, the furanose rings, the phosphate groups, etc.) The geometry of the polymer is described by the set of torsional angles which define the orientations about these single bonds. A very simple example is ethane which can be considered to consist of two rigid methyl groups connected by a C-C bond about which rotation can take place. The geometry of ethane (i.e. hydrogens staggered, eclipsed, etc.) is described by the torsional angle for the C-C bond.

This formalism requires that the identity of all atoms and their positions in the small, rigid groups be known. In the case of simple organic high polymers this is usually a trivial task. In the case of most biopolymers it is not. However, recently investigators have determined the sequence of many biopolymers (especially proteins and nucleic acids) for which the secondary, tertiary and quarternary structures are unknown.

In the calculations the geometry of these rigid groups invariably have been taken from the crystal structures of small model compounds (e.g. amino acids or nucleosides).

It would be better if we could get these data from the structures of proteins or nucleic acids. This is a moot point for nucleic acids where attempts to crystalize RNAs have been unsuccessful.

The possibility that bond lengths or bond angles change (from one conformation to another) is usually neglected in these treatments. Changes in conformation caused by distortions of the supposedly rigid groups are small compared with the changes resulting from rotations about single bonds[10]. These distortions will be greatest when there are large energy gradients. Because these gradients are usually associated with strong repulsive interactions which result in high energy conformations and we are interested mostly in low energy conformations, these effects can be neglected in most of our work.

The one remaining facet of the formalism of the calculation is the choice of potential. Theoretical calculations on simple molecules (e.g. ethane) have shown that the barrier to rotation about single bonds can be accurately approximated by adding the pairwise interactions of the atoms on either side of the bond[37]. In order to simplify calculations, workers studying polypeptide structures usually express these energies as explicit functions of the torsional angles. Potentials of the form $A\cos(n\theta) + B\cos(2n\theta) + C\cos(3n\theta) + \ldots$ (where the bond has n-fold rotational symmetry) are commonly used. The coefficients can be calculated by computing the potential as a function of torsional angle, (assuming that the potential arises

from pairwise charge-charge, charge-polarizability, London dispersion and steric repulsion interactions of the atoms close to either end of the bond) then fitting these numerical values to the appropriate Fourier series. In order to check this approximation of the potential energy, calculations of the energy barriers for internal rotation in the cases of simple molecules (e.g. acetone) were made[170]. Agreement with barriers experimentally measured (usually by microwave spectroscopy) was quite good.

This type of potential does not take into account interactions between atoms which are not near-neighbors in the primary structure of the molecule. We know that polypeptide chains (and probably polynucleotide chains as well) are frequently folded in such a way that groups not close to one another in the primary sequence are close to one another in the three dimensional structure. In order to include important interactions between atoms far from one another in the primary sequence, long-range electrostatic interactions (frequently expressed as dipole-dipole interactions) and short range interactions like hydrogen bonds and steric repulsions, are added to the potential.

The simplest approach to predicting the conformation of biopolymers on the microscopic level is one in which structures with unfavorable short interatomic contacts are discarded. This technique has been applied to both polynucleotides[166] and polypeptides[135,156-158]. We will discuss the results of the polynculeotide work later, comparing them with our results. This approach leads to two problems.

First, there is no way of determining the energy of those conformations allowed. Secondly, large repulsive interactions do not arise until the atoms are well within minimum contact distances found in crystals[10]. This fact is manifest in the prediction that certain structures found in X-ray studies of protein crystals are forbidden by hard core contact criteria[157]. If the hard core radii are reduced enough to insure that no reasonable structures are excluded, then the method fails to eliminate enough structures which have high energies and looses most of its attractiveness. In order to avoid this difficulty it is better to use a "soft" potential such as $Ae^{-br}$ or $Ar^{-12}$. This way closer contact distances will have progressively higher energies unless there is an important interaction (such as a hydrogen bond) which lowers the energy considerably and makes the conformation feasible.

Present treatments of calculations of nucleic acid structures which consider the ribose-phosphate backbone are incomplete. An early attempt to predict the structure of the acid form of poly A on the basis of assigning orientation angles for rotations about single bonds[130] (as found in small model compounds) was ahead of its time in the approach to this problem. However later work has shown this treatment to have been incorrect in its execution[161]. Other investigations have considered which angular orientations are "forbidden" and "allowed" on the basis of hard core repulsions[166]. They do not include a systematic treatment of attractive or repulsive interactions other than the hard

core. A number of treatments have been published which treat "stacked"[36,48,153] and coplanar hydrogen-bonded[134,146,154] bases. Their results indicate that these aggregates are stable, relative to the separated bases and enable comparisons to be made of their relative stabilities. Most of this work is directed towards the problems of the stabilities of multi-strand polynucleotides. Our interest is largely in single-strand oligomers, dinucleoside phosphates in particular. On the other hand, calculations of conformations of oligo- and polypeptides are more pertinent, as they include reasonably successful treatments which do include extensive treatments of backbone conformations.

Calculations of the conformations of oligopeptides and polypeptides based on the microscopic approach have taken two basic forms. Both are essentially calculations of the two dimensional energy surface (There can be rotation about the $C^{\alpha}$-C' and N-$C^{\alpha}$ bonds of a polypeptide backbone.), sometimes including the effects of varying the orientation of the amino acid side-chain. One approach calculates the average properties of coiled polypeptides[17-19,129,168], and has good agreement with experimental results. However these calculations may not be successful when applied to rigid polypeptide structures because the statistical nature of the test may hide inaccuracies which would become important in a situation where the properties of large numbers of conformations are not averaged.

The other approach treats rigid structures. Usually calculations are restricted to small molecules such as dipeptides[20,64] or else homopolymers where the symmetry of the polymer simplifies the calculation[45,115,136,171]. Tests of these calculations have come from comparisons with conformations found either in synthetic polypeptides or in crystal structures of proteins (i.e. myoglobin or lysozyme). In one respect the agreement is good; structures found experimentally are generally very close to minima of the potential energy surfaces. It seems certain that the locations of these minima are accurate. However, many structures are not near the lowest appropriate minimum, showing that the accuracy of the calculated energy differences must be improved. Because these minima are not close to one another, one wrong choice for the orientation angles for just one amino acid residue in a reasonably long polypeptide chain could have disastrous effects. This mistake might radically change the direction of propagation of the polypeptide chain.

Calculation of the potential energy surface is difficult for molecules which are long and do not have the symmetry of a polypeptide composed of only one amino acid. For a small protein, with 200 amino acids, 3 degrees of freedom for each amino acid and 10 orientations for each degree of freedom, we would have to try $10^{600}$ different conformations. Of course, most of these conformations

contain redundancies. (We expect orientations of the side chains of similar amino acids to be similar, etc.) Even if this number could be reduced by a factor of $10^{500}$ there would still be $10^{100}$ conformations, an astronomical number.

The two microscopic calculations of oligopeptides of moderate length (about 10 amino acids) take different approaches in getting around this problem. The first calculation was for Gramacidin S, a cyclic decapeptide with a repeated pentapeptide sequence[202]. The repeated sequence and the fact that the polypeptide chain had to return to the place from which it started, greatly simplified the calculation. The structure proposed for Gramacidin-S is compatible with preliminary X-ray diffraction studies. It will be interesting to see how this comparison fares when more detailed X-ray data becomes available.

The other study was on the S-peptide of bovine pancreatic ribonuclease A[65]. In this case the oligopeptide was placed into reasonable conformations (such as the α-helix), and minimization techniques were tried to see if the oligopeptide would relax into the lowest energy conformation. If, starting from different conformations, the same end point was found, this would almost certainly be the unique stable conformation. Unfortunately a potential energy surface as complicated as this (i.e. with as many dimensions) is bound to have a large number of local energy minima. The minimization procedure was rapidly trapped in local energy minima, never having a chance to go far enough to find the true minimum.

The macroscopic technique tries to get around these difficulties by using additional information about the geometry of the polymer. Often amino acids (not close to one another along the primary sequence) are known to be at the active site of an enzyme. Applying this constraint would require folding the polypeptide chain to bring the amino acids associated with the active site close to one another. In certain cases other amino acids not close to one another in the primary sequence are known to interact. An example of this is the case of the interactions of the buried tyrosines of bovine pancreatic ribonuclease[59,60,111,218]. In myoglobin and lysozyme the outside of the molecule is largely hydrophilic while the inside is almost entirely hydrophobic[141,142]. (Perhaps this rule applies to most other proteins as well.) Using these constraints, a model of the biopolymer is constructed. Then a pseudo-potential of a very simple form (e.g. an $r^{-1}$ potential attracting hydrophobic groups to one another) is applied which can either attract or repel certain parts of the molecule. A search for the energy minima (using steepest descent techniques, for example) then yields the most likely answers.

Two groups have been actively working on the problem of trying to predict the structure of proteins by the macroscopic approach. A review by Levinthal[107] summarizes the approaches which can be taken to such a problem. Up to this time his group has not published predictions of protein structures which could be compared with the results of X-ray crystallographic studies. On the other hand

Hammes and Scheraga[71] have proposed a structure for bovine pancreatic ribonuclease A. The data used in developing their model were the positions of the four disulfide bonds, the interactions of the three buried tyrosines and the fact that the other three tyrosines had to be at the surface, the proximity of the amino acids known to participate at the active site and the over all dimensions of the molecule which were determined by preliminary X-ray diffraction studies. Shortly after this paper was published, the solution of the X-ray structure of ribonuclease at the 2Å level was announced[93]. The model proposed on the basis of chemical data is quite similar to the structure determined by X-ray diffraction techniques. However, substantial differences between the two structures do exist.

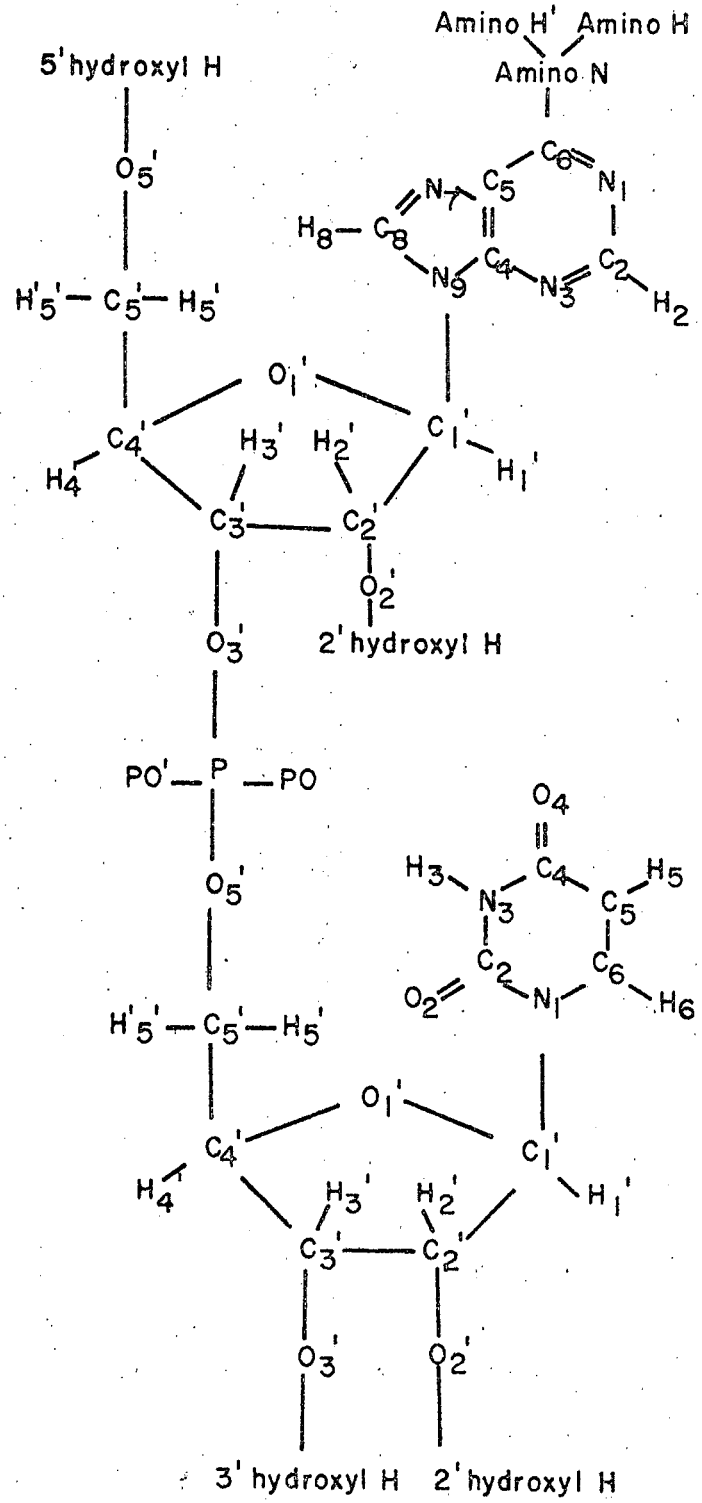
There is no reason why the same approach used for ribonuclease could not be used for a transfer RNA. Susceptibility to enzymatic attack and the reactivity of bases to modifying reagents indicate which bases are exposed and which are protected (presumably in multiple-strand hydrogen-bonded structures). Small angle X-ray diffraction studies have already determined the over-all dimensions of transfer RNAs[102,200]. It was reasoning with this type of information which led to the proposal of the cloverleaf model for tRNA[61]. Considerations of more chemical data, along with the use of pseudo-potentials to minimize the energy with respect to the finer points of the structure can lead to more detailed models.

All these calculations have been very empirical. The number of parameters which we must use will be staggering, considering the approximate nature of all of them. Until proven otherwise the results must only be considered as suggestions. Another point common to almost all these calculations is one so obvious that it is frequently missed by a person who has never done this type of calculation; the fact that the calculations are very long. The difference between a successful calculation and no calculation at all is the ability of the investigator to cut corners. It should be remembered, that in his eagerness to save time it is very easy to discard a conformation which may be the one for which the investigator is searching. The value of caution should not be underestimated.

METHODS

Since we are most interested in the conformations of small molecules the microscopic approach would seem most applicable. First the potential and the geometries of the rigid segments of the dinucleoside phosphates must be chosen. In the case of polypeptides, Pauling, Corey and their coworkers systematically determined the crystal structures of a large number of amino acids and oligopeptides[140]. This extensive survey showed the similarities and differences among amino acids. It was culminated in their prediction of hydrogen-bonded helical and sheet structures for polypeptides. Unfortunately, no extensive survey of analogous

# ATOMS OF A DINUCLEOSIDE PHOSPHATE




XBL6711-5726

Figure 28 Nomenclature used in denoting the atoms of a dinucleoside phosphate.

nucleic acid structures has been undertaken. This is probably because a nucleotide is considerably larger than an amino acid and the synthesis of oligonucleotides is considerably more difficult than the synthesis of oligopeptides. This makes the choices of geometries of the groups composing nucleic acids (especially the sugars) arbitrary, to a large extent.

The atomic coordinates of the purine and pyrimidine bases were taken to be those proposed by Spencer[177]. They are listed in Table 9. The entire base is planar, including the $C_1'$ of the ribose. More recent experimental results show that this assumption is good for the atoms of the rings while extranuclear substituents occasionally can deviate significantly from the plane of the ring (up to 0.033Å for adenine $N_6$ and 0.3Å for $C_1'$)[42]. Much of this may be caused by crystal packing forces, rather than intramolecular effects which would be present in isolated molecules.

The coordinates of uracil were the same as those proposed for thymine except for $H_5$ which is substituted for the methyl group. The coordinates of $H_5$ were determined by having the $C_5$-$H_5$ bond oriented in the same direction as the $C_5$-methyl of thymine, with a bond length equal to the other $C(sp^2)$-H bond lengths in cytosine and thymine.

For pseudouracil, the glycosidic bond (now a C-C bond) was assumed to be 1.54Å long (the C-$CH_3$ distance in thymine)[82,193] and in the same direction as the $C_5$-$CH_3$ bond. The $N_1$-$H_1$ bond was assumed to be in the direction of the glycosidic bond of thymine, with a bond length equal

TABLE 9

ATOMIC COORDINATES OF PURINE AND PYRIMIDINE BASES

Adenine

| Atom | X | Y |
|---|---|---|
| $N_1$ | -2.791Å | 4.407Å |
| $C_2$ | -3.201 | 3.134 |
| $N_3$ | -2.391 | 2.078 |
| $C_4$ | -1.079 | 2.298 |
| $C_5$ | -0.604 | 3.583 |
| $C_6$ | -1.500 | 4.633 |
| $N_7$ | 0.763 | 3.598 |
| $C_8$ | 1.055 | 2.280 |
| $N_9$ | 0.000 | 1.470 |
| $H_2$ | -4.284 | 2.943 |
| Amino N | -1.041 | 5.892 |
| Amino H | -1.773 | 6.574 |
| Amino H' | -0.067 | 6.117 |
| $H_8$ | 2.089 | 1.904 |

Guanine

| | | |
|---|---|---|
| $N_1$ | -2.799 | 4.348 |
| $C_2$ | -3.205 | 3.051 |
| $N_3$ | -2.378 | 2.010 |
| $C_4$ | -1.079 | 2.298 |
| $C_5$ | -0.604 | 3.583 |
| $C_6$ | -1.462 | 4.702 |
| $N_7$ | 0.763 | 3.598 |
| $C_8$ | 1.055 | 2.280 |
| $N_9$ | 0.000 | 1.470 |
| Amino N | -4.523 | 2.807 |
| Amino H | -4.904 | 1.883 |
| Amino H' | -5.191 | 3.532 |
| $H_1$ | -3.556 | 5.145 |
| $O^1$ | -1.045 | 5.848 |
| $H_8$ | 2.089 | 1.904 |

Cytosine

| Atom | X | Y |
|---|---|---|
| $N_1$ | 0.000Å | 1.470Å |
| $C_2$ | -1.207 | 2.139 |
| $N_3$ | -1.231 | 3.489 |
| $C_4$ | -0.070 | 4.132 |
| $C_5$ | 1.157 | 3.504 |
| $C_6$ | 1.181 | 2.125 |
| $O^6$ | -2.253 | 1.511 |
| Amino N | -0.094 | 5.472 |
| Amino H | -0.942 | 6.002 |
| Amino H' | 0.735 | 6.031 |
| $H_5$ | 2.109 | 4.054 |
| $H_6$ | 2.133 | 1.575 |

Uracil

| | | |
|---|---|---|
| $N_1$ | 0.000 | 1.470 |
| $C_2$ | -1.207 | 2.139 |
| $N_3$ | -1.159 | 3.518 |
| $C_4$ | 0.010 | 4.251 |
| $C_5$ | 1.205 | 3.504 |
| $C_6$ | 1.181 | 2.125 |
| $O_2$ | -2.269 | 1.538 |
| $H_3$ | -2.051 | 3.954 |
| $O_4$ | 0.010 | 5.471 |
| $H_5$ | 2.176 | 4.021 |
| $H_6$ | 2.124 | 1.558 |

Pseudouracil

| | | |
|---|---|---|
| $N_1$ | -1.231 | 3.559 |
| $C_2$ | -0.073 | 4.311 |
| $N_3$ | 1.122 | 3.620 |
| $C_4$ | 1.220 | 2.244 |
| $C_5$ | 0.000 | 1.540 |
| $C_6$ | -1.207 | 2.209 |
| $H_1$ | -2.184 | 4.066 |
| $O_2$ | -1.052 | 5.530 |
| $H_3$ | 1.925 | 4.203 |
| $O_4$ | 2.120 | 1.766 |
| $H_6$ | -2.150 | 1.643 |

to the other N-H bond of thymine (1.08Å)

The coordinates of the bases were calculated with a desk calculator using the bond lengths and bond angles given. The results were checked by determining the position of one atom on each ring by going around the ring in either direction. The deviations found were ±0.01Å or less in all cases. This was to be expected as a result of the lack of self-consistency (round-off errors) in the data from the literature. Atomic coordinates determined graphically agreed with our results.

The ribose conformation was not as easy to choose. X-ray crystallographic studies have shown that four of the five members of the furanose ring are very close to being coplanar, with the fifth member of the ring about 0.4 or 0.5Å out of the plane[180,183]. The conformation of the ring is described by noting which atom is out of the plane and whether it is on the same side of the plane as $C_5'$ (endo) or on the opposite side (exo). The 2' endo conformation has been found most frequently, with the 3' endo conformation also common.

It has been proposed that this puckering originates in repulsive interactions between substituents attached to to the five-membered ring which would be eclipsed if the ring were planar[180]. In the case of cyclopentane, the molecule gains about 4 kcal/mole by puckering[97,145]. The energy differences between different puckered forms of the furanose ring may be smaller than 4 kcal per mole. NMR measurements which indicate that the ribose ring of nucleosides and nucleotides is in the 2' endo conformation

in aqueous solution. It is based on a theory which calculates the spin-spin splitting of hydrogens attached to neighboring carbons as a function of the dihedral angles between the two C-H bonds[92]. Splittings of $H_1'$ by $H_2'$ (which depend on the dihedral angle between C-H bonds and therefore, the nature of the furanose puckering) have been measured[85,86]. The purine nucleosides and nucleotides seem to be in the 2' endo conformation. The agreement of the results for pyrimidine nucleosides and nucleotides with the theory is not as good; but the splitting is closer to that expected for the 2' endo conformation than any other reasonable structure. As a control, the splitting of the $H_1'$ resonance of 3'-5' cyclic adenylic acid (where the phosphodiester bridge forces the furanose ring into a 3' endo conformation) was measured and found to be what would be expected for a 3' endo conformation. We feel that the furanose ring is probably puckered with $C_2'$ puckered approximately 0.45Å from the plane of the other four atoms and on the same side as the $C_5'$. As a precaution, we have also calculated the energies of conformations of compounds where the furanose ring had the 3' endo conformation.

The coordinates for 2' endo ribose were taken from the structure of Cp.[182] The 3' hydroxyl group was aligned along the $O_3'$-P bond of the nucleotide with an O-H bond length of 1.00Å. The corresponding deoxyribose coordinates were the same with the exception of $H'_2'$ which was situated along the $C_2'$-$O_2'$ bond of the ribose with a bond length

TABLE 10 PART 1

ATOMIC COORDINATES

2' ENDO RIBOSE[182]

| Atom | X | Y | Z |
|---|---|---|---|
| $C_1'$ | 0.000 | 0.000 | 0.000 |
| $H_1'$ | -0.510 | -0.161 | 0.740 |
| $O_1'$ | 1.350 | -0.432 | 0.000 |
| $C_2'$ | -0.623 | -0.588 | -1.248 |
| $H_2'$ | -0.262 | 0.083 | -1.895 |
| $O_2'$ | -2.025 | -0.567 | -1.264 |
| 2' Hydroxyl H | -2.610 | -1.090 | -0.593 |
| $C_3'$ | 0.080 | -1.949 | -1.316 |
| $H_3'$ | 0.017 | -2.338 | -2.021 |
| $C_4'$ | 1.509 | -1.595 | -0.864 |
| $H_4'$ | 1.816 | -2.457 | -0.379 |
| $C_5'$ | 2.487 | -1.287 | -1.973 |
| $H_5'$ | 3.374 | -0.735 | -1.408 |
| $H'_5'$ | 2.988 | -2.214 | -2.361 |
| $O_5'$ | 1.947 | -0.437 | -2.999 |
| 5' Hydroxyl H | 1.883 | -1.151 | -3.697 |
| P | -0.441 | -4.410 | -0.467 |
| Phosphate O (P-O) | -1.390 | -4.070 | -1.170 |
| Phosphate O' (P-O) | -1.020 | -4.860 | 0.898 |
| Phosphate O (P-O-H) | 0.967 | -4.820 | -0.680 |
| $N_1$ | 0.000 | 1.470 | 0.000 |

TABLE 10 PART 2

ATOMIC COORDINATES

3' ENDO RIBOSE[99]

| Atom | X | Y | Z |
|---|---|---|---|
| $C_1'$ | 0.000 | 0.000 | 0.000 |
| $H_1'$ | - 0.758 | -0.199 | 0.801 |
| $O_1'$ | 1.380 | -0.428 | 0.000 |
| $C_2'$ | -0.633 | -0.581 | -1.241 |
| $H_2'$ | -1.394 | 0.011 | -1.725 |
| $O_2'$ | -1.064 | -1.912 | -0.909 |
| 2' Hydroxyl H | -1.625 | -2.236 | -1.579 |
| $C_3'$ | 0.570 | -0.642 | -2.207 |
| $H_3'$ | 0.776 | 0.306 | -2.697 |
| $O_3'$ | 0.373 | -1.406 | -3.363 |
| 3' Hydroxyl H | 0.460 | -2.161 | -3.077 |
| $C_4'$ | 1.640 | -1.155 | -1.258 |
| $H_4'$ | 1.502 | -2.086 | -0.964 |
| $C_5'$ | 3.109 | -0.954 | -1.618 |
| $H_5'$ | 3.533 | -1.645 | -2.310 |
| $O_5'$ | 3.245 | 0.393 | -2.203 |
| P | 4.753 | 0.811 | -2.583 |
| Phosphate O | 5.376 | -0.306 | -3.394 |
| Phosphate O' | 4.652 | 2.166 | -3.207 |
| Phosphate O'' | 5.519 | 0.989 | -1.229 |
| $N_1$ | 0.000 | 1.470 | 0.000 |

TABLE 10 PART 3

ATOMIC COORDINATES

3' ENDO DEOXYRIBOSE[193]

| Atom | X | Y | Z |
|---|---|---|---|
| $C_1'$ | 0.000 | 0.000 | 0.000 |
| $H_1'$ | -0.402 | -0.533 | 0.958 |
| $O_1'$ | 1.372 | -0.429 | 0.000 |
| $C_2'$ | -0.676 | -0.645 | -1.207 |
| $H_2'$ | -1.281 | -1.468 | -1.108 |
| $H'_2'$ | -1.347 | -0.163 | -1.863 |
| $C_3'$ | 0.519 | -1.075 | -2.096 |
| $H_3'$ | 0.866 | -0.278 | -2.799 |
| $O_3'$ | -0.241 | -2.156 | -2.956 |
| 3' Hydroxyl H | 0.056 | -2.999 | -2.401 |
| $C_4'$ | 1.567 | -1.370 | -1.076 |
| $H_4'$ | 1.198 | -2.364 | -0.703 |
| $C_5'$ | 3.021 | -1.263 | -1.524 |
| $H_5'$ | 3.712 | -1.637 | -0.797 |
| $H'_5'$ | 3.276 | -1.805 | -2.388 |
| $O_5'$ | 3.273 | 0.090 | -2.043 |
| 5' Hydroxyl H | 4.058 | 0.232 | -2.726 |
| $N_1$ | 0.000 | 1.470 | 0.000 |

of 1.018Å. 3' endo ribose and deoxyribose coordinates were taken from the structures of pA[99] and calcium thymidylate[193]. Location of hydrogen atoms are quite difficult in crystals of molecule as complex as nucleotides. In the case of pA[99] one of the $H_5'$ could not be located unambiguously. Since the $H_5'$ is far from the purine or pyrimidine base, and has a small charge and small polarizability, we neglected all interactions involving this atom. The coordinates of the phosphate group were taken from Cp[182]. In forming CpC the structure of Cp was used. The $C_5'$-$O_5'$ bond was formed by aligning the $C_5'$-$O_5'$ bond along the same line as the phosphate O-H of Cp, with a $C_5'$-$O_5'$ bond length equal to the $C_3'$-$O_3'$ bond length of Cp.

The atomic coordinates for 2' endo ribose were calculated by hand from the published coordinates (in terms of unit cell dimensions). They were checked by computing at least one interatomic distance for each atom and then comparing them with the values listed in the paper. Two minor typographical errors were discovered[88]. The coordinate x/a for $H_4'$ should be -0.01349 instead of +0.01349 and the length of the $C_2'$-$H_2'$ bond should be 0.999Å instead of 1.999Å. Other atomic coordinates were calculated by computer programs. The coordinates of 2' endo ribose calculated in this way agreed with those calculated by hand. All coordinates are listed in Table 10.

In addition to deciding on what geometries we use for the rigid groups which make up oligonucleotides we must define conventions for describing rotations about single bonds. In contrast to the situation with polypeptides,

TABLE 11

| Bond | Atoms Defining Standard Orientation | Sense of Rotation |
|---|---|---|
| Glycosidic | $O_1'$, $C_8$ or $C_6$ | Clockwise |
| $C_2'$-$O_2'$ | $C_1'$, 2' hydroxyl H | Counter-clockwise |
| $C_3'$-$O_3'$ | $C_4'$, P | Counter-clockwise |
| $O_3'$-P | $C_3'$, $O_5'$ | Counter-clockwise |
| P-$O_5'$ | $O_3'$, $C_5'$ | Counter-clockwise |
| $O_5'$-$C_5'$ | $C_4'$, P | Counter-clockwise |
| $C_5'$-$C_4'$ | $O_1'$, $O_5'$ | Counter-clockwise |

where there is a widely accepted convention[52], we must develop a way of describing rotations about single bonds. The orientation about the glycosidic bond is defined in accordance with the accepted convention of Donohue and Trueblood[59]. The two regions with no strong steric repulsions are the anti conformation in which $H_6$ of pyrimidines of $H_8$ of purines is close to $O_1'$ ($\phi_{CN} \sim -20°$) and the syn conformation in which $H_6$ of pyrimidines or $H_8$ of purines is far from $O_1'$ ($\phi_{CN} \sim 140°$). Orientations about other single bonds are defined in a slightly different way. Two atoms define the ends of each bond. Each of these atoms has an atom bonded to it which defines the standard (0°) orientation. When all four atoms are in a cis-coplanar configuration the bond is in its standard orientation. Looking down the bond, if the group closest to the viewer is rotated counter-clockwise then the angle of orientation increases. (According to this scheme, Donohue and Trueblood define increasing angles for rotations about the glycosidic bond in the opposite sense.) The atoms defining the standard orientation are listed in Table 11.

The other choice we have to make is the form of the potential. Theoretical calculations on simple molecules (e.g. ethane) have shown that the barrier to rotation about single bonds can be predicted on the basis of pairwise interactions of atoms on either side of the bond[37]. We therefore chose a potential consisting of pairwise interactions of atoms on either side of the bond. To simplify

our calculations we included only those interactions which change with internal rotation. Four types of interactions were considered: charge-charge, charge-polarizability, London dispersion and steric repulsion interactions. The first two terms are reasonably straightforward. For uncharged polypeptides it is convenient to express charge distributions in terms of point dipoles. In the case of oligonucleotides this is not a good approximation. There is a charged phosphate group which is not a dipole in any approximation. Also, the charge distributions in the rigid groups composing nucleotides are considerably more complex then those found in amino acids. This is compounded by the fact that we will have to consider conformations in which the distances between groups will be smaller than the dimensions of the groups themselves. This situation will introduce serious errors even with far simpler charge distributions. The London dispersion interaction ($r^{-6}$ attraction) is a drastic approximation. It is derived for the case where the interacting groups are far from one another[120]. There are appreciable deviations from the $r^{-6}$ distance dependence when atoms come close. In addition, the simple assumption that the polarizability is all located at a point (the nucleus) and is isotropic may cause difficulties, especially because we must consider aromatic electronic systems (where the polarizable delocalized electron distribution has a node at the nucleus). However, it has been found that substitution of isotropic polarizabilities for anisotropic ones does not appreciably

change the calculated London dispersion interactions for "stacked" bases[36]. Since isotropic polarizabilities greatly facilitate the calculation, we have used them. We make no claims for the accuracy of a London dispersion potential at such short range. We use it to represent a short range attractive force which includes major contributions from London-type dispersion interactions. Pitzer's expression for the London dispersion is used[144].

The remaining term was the steric repulsion. At first it might seem that a hard core repulsion (one in which an interatomic distance smaller than a specified minimum contact distance causes the potential to become infinite) is the best way to treat steric repulsions. We feel that this is not the case for two reasons. Interatomic distances considerably closer than the intermolecular Van der Waals distances can exist without very large repulsions[10]. The second reason is that a hard core potential is not differentiable. (It is discontinuous at its hard core contacts.) Mathematical operations are almost invariably easier if the potential is differentiable. The best situation is where the potential is analytic. Two types of analytic repulsive potentials have been used: an exponential potential of the type $Ae^{-br}$ and the type $Ar^{-b}$ where b is usually chosen to be 12. The exponential repulsion has the disadvantage that at very small interatomic distances the dominating term of the potential is the attractive London dispersion term. We know that this is physically unreasonable. We have therefore chosen the $Ar^{-12}$ form for the steric repulsive potential

with the constant A chosen so that the $r^{-6}$ and $r^{-12}$ potentials when combined have a minimum at the intermolecular Van der Waals distance and increase monotonically for smaller interatomic distances.

No explicit hydrogen bond potential has been included. However, hydrogen bonds are implicity included in the electrostatic terms. Nash and Bradley[134] found very reasonable hydrogen bond geometries for purine and pyrimidine base-pairs using only charge-charge interactions.

Finally, these calculations are done for molecules in a vacuum. There are no solvent effects and a unit dielectric constant is assumed. Attempts to calculate solvent interactions in the framework of the microscopic model have been made for oligopeptides[65]. Unfortunately the molecule being considered was sufficiently complicated that no lowest energy conformations had been reached. Since no comparisons with experiment can be made at this time it is too early to see whether their method for including solvent effects will be successful. The experimental studies of dinucleoside phosphates and other nucleic acid materials have not enabled us to determine the manner in which water affects the conformations of these compounds. Perhaps an empirical approach correlating the solvent effects with an effective temperature or set of dielectric constants can be used.

The energy of interaction between atoms i and j is therefore:

$$E_{ij} = 3.318\text{x}10^5 \frac{\rho_i \rho_j}{r_{ij}} \qquad \text{(charge-charge)}$$

$$- 1.659\text{x}10^5 \frac{(\rho_i^2 \alpha_j + \rho_j^2 \alpha_i)}{r_{ij}^4} \qquad \text{(charge-polarizability)}$$

$$- 3.650\text{x}10^5 \frac{\alpha_i \alpha_j}{r_{ij}^6 \left[\left(\frac{\alpha_i}{N_{eff_i}}\right)^{\frac{1}{2}} + \left(\frac{\alpha_j}{N_{eff_j}}\right)^{\frac{1}{2}}\right]} \qquad \text{(London dispersion)}$$

$$+ 1.825\text{x}10^5 \frac{r_{VW}^6 \alpha_i \alpha_j}{r_{ij}^{12} \left[\left(\frac{\alpha_i}{N_{eff_i}}\right)^{\frac{1}{2}} + \left(\frac{\alpha_j}{N_{eff_j}}\right)^{\frac{1}{2}}\right]} \qquad \text{(steric repulsion)}$$

where $E_{ij}$ is the energy of interaction (in calories per mole), $\rho_i$ and $\rho_j$ are the charges of atoms i and j (in electrons), $r_{ij}$ is the distance between atoms i and j (in Å), $\alpha_i$ and $\alpha_j$ are the polarizabilities of atoms i and j (in $Å^3$), $N_{eff_i}$ and $N_{eff_j}$ are the "effective number of electrons of atoms i and j (see ref. 144) and $r_{VW}$ is the sum of the Van der Waals radii of atoms i and j.

In considering the four terms which contribute to the potential energy of interaction between two atoms the question arises: Can one or a few interatomic interactions be singled out as being critical in stabilizing or prohibiting a structure? The answer to this question is usually no. The one exception is the steric repulsion ($r^{-12}$) term. At very short interatomic distances this term dominates all others. All other interactions are generally smaller than a few kcal per mole (except in the case of absurdly small interatomic contacts in which case the $r^{-12}$ term dominates anyway). Single interatomic interactions are frequently

smaller than 1 cal. per mole.

Experimental studies of the thermal disordering of dinucleoside phosphates indicate that we are looking for energy differences between important conformations of approximately 5 to 10 kcal per mole. Because this difference originates in changes in a large number of terms, we usually will not be able to single out one or a few interactions which are largely responsible for the energy differences between conformations.

In the attractive region of the 6-12 potential this interaction rarely exceeds 1 kcal per mole. The charge-charge and charge-polarizability interactions frequently do. It should be noted that charge-polarizability interactions are always attractive while charge-charge interactions may be either attractive or repulsive.

The expression we have chosen to use for the interatomic interaction potential requires the knowledge of atomic charges, polarizabilities, "effective number of electrons" and Van der Waals radii. The distributions of $\pi$ electrons were calculated using semi-empirical self-consistent molecular orbital wave functions[9,150]. Wave functions of this type (when combined with the $\sigma$ electron treatment which was used in our work) predict dipole moments of nitrogen-containing aromatic compounds (pyridine, pyrimidine, pyrrole, etc.) which agree well with experimentally measured values. This indicates that the charge distributions are probably good. The $\sigma$ charge

distributions were calculated using the method of Del Re[44]. Interactions between $\pi$ and $\sigma$ electrons are introduced in the self-consistency criterion of the molecular orbital wave functions. The charge distribution in the phosphate group was taken to be that calculated for the diethyl phosphate anion by a self-consistent molecular orbital method[38]. The phosphate group is electronegative and is assumed to withdraw 0.16405 electron from $C_3'$ and $C_5'$. The charge distributions are listed in Table 12.

A cursory examination of the charge distributions shows how bad the dipole approximation is. If there had been relatively few charged atoms in each group or the positively charged atoms tended to be on one side of the rigid group and the negatively charged atoms on the other, then the approximation might have been satisfactory. The charge distributions we calculate does not meet either criterion. In groups which are essentially neutral there are a number of highly charged atoms which are distributed in a rather irregular way, making it useless (for this calculation) to try to approximate the distribution as a point dipole.

The charge distributions used in our calculations were very similar to those used by other investigators. A comparison of our charge distributions with those used by Bradley and coworkers[13,134] is presented in Table 13.

The isotropic polarizabilities of all atoms except phorphorus are taken from Ketelaar's monograph[96]. These are the values used by almost all other investigators.

TABLE 12 PART 1

CHARGE DISTRIBUTIONS

| Atom | π Charge | σ Charge (Ribonucleoside) | Total Charge (Ribonucleoside) | σ Charge (Deoxyribonucleoside) | Total Charge (Deoxyribonucleoside) |
|---|---|---|---|---|---|
| **Adenosine** | | | | | |
| $N_1$ | -0.265 | -0.127 | -0.392 | -0.127 | -0.392 |
| $C_2$ | 0.164 | 0.085 | 0.249 | 0.085 | 0.249 |
| $N_3$ | -0.273 | -0.128 | -0.401 | -0.128 | -0.401 |
| $C_4$ | 0.029 | 0.132 | 0.161 | 0.132 | 0.161 |
| $C_5$ | -0.071 | 0.086 | 0.015 | 0.086 | 0.015 |
| $C_6$ | 0.146 | 0.149 | 0.295 | 0.149 | 0.295 |
| $N_7$ | -0.323 | -0.133 | -0.456 | -0.133 | -0.456 |
| $C_8$ | 0.063 | 0.091 | 0.154 | 0.091 | 0.154 |
| $N_9$ | 0.368 | -0.198 | 0.169 | -0.200 | 0.168 |
| Amino N | 0.161 | -0.531 | -0.369 | -0.531 | -0.369 |
| Amino H | | 0.217 | 0.218 | 0.218 | 0.218 |
| $H_2$ | | 0.042 | 0.042 | 0.042 | 0.042 |
| $H_8$ | | 0.043 | 0.043 | 0.043 | 0.043 |
| **Guanosine** | | | | | |
| $N_1$ | 0.293 | -0.352 | -0.058 | -0.352 | -0.058 |
| $C_2$ | 0.117 | 0.215 | 0.332 | 0.215 | 0.332 |
| $N_3$ | -0.386 | -0.120 | -0.505 | -0.120 | -0.505 |
| $C_4$ | -0.031 | 0.133 | 0.102 | 0.133 | 0.102 |
| $C_5$ | -0.068 | 0.091 | 0.022 | 0.091 | 0.022 |
| $C_6$ | 0.215 | 0.198 | 0.413 | 0.198 | 0.413 |
| $N_7$ | -0.273 | -0.132 | -0.405 | -0.132 | -0.405 |
| $C_8$ | 0.027 | 0.091 | 0.118 | 0.091 | 0.118 |
| $N_9$ | 0.386 | -0.198 | 0.189 | -0.199 | 0.187 |
| Amino N | 0.148 | -0.526 | -0.378 | -0.526 | -0.378 |
| Amino H | | 0.219 | 0.219 | 0.219 | 0.219 |
| $H_3$ | | 0.202 | 0.202 | 0.202 | 0.202 |
| $O_6$ | -0.428 | -0.136 | -0.563 | -0.136 | -0.563 |
| $H_8$ | | 0.043 | 0.043 | 0.043 | 0.043 |

TABLE 12 PART 2

CHARGE DISTRIBUTIONS

| Atom | π Charge | σ Charge (Ribonucleoside) | Total Charge (Ribonucleoside) | σ Charge (Deoxyribonucleoside) | Total Charge (Deoxyribonucleoside) |
|---|---|---|---|---|---|
| | | Uridine | | | |
| $N_1$ | 0.270 | -0.193 | 0.078 | -0.194 | 0.076 |
| $C_2$ | 0.194 | 0.247 | 0.441 | 0.248 | 0.441 |
| $N_3$ | 0.260 | -0.257 | 0.002 | -0.257 | 0.002 |
| $C_4$ | 0.185 | 0.025 | 0.210 | 0.025 | 0.210 |
| $C_5$ | -0.089 | 0.034 | -0.055 | 0.034 | -0.055 |
| $C_6$ | 0.015 | 0.041 | 0.057 | 0.041 | 0.056 |
| $O_2$ | -0.451 | -0.129 | -0.581 | -0.129 | -0.581 |
| $H_3$ | | 0.188 | 0.188 | 0.188 | 0.188 |
| $O_4$ | -0.384 | -0.085 | -0.469 | -0.085 | -0.469 |
| $H_5$ | | 0.037 | 0.037 | 0.037 | 0.037 |
| $H_6$ | | 0.038 | 0.038 | 0.038 | 0.038 |
| | | Cytidine | | | |
| $N_1$ | 0.310 | -0.194 | 0.116 | -0.195 | 0.115 |
| $C_2$ | 0.207 | 0.241 | 0.448 | 0.241 | 0.448 |
| $N_3$ | -0.297 | -0.118 | -0.415 | -0.118 | -0.415 |
| $C_4$ | 0.128 | 0.130 | 0.258 | 0.130 | 0.258 |
| $C_5$ | -0.134 | -0.011 | -0.146 | -0.011 | -0.146 |
| $C_6$ | 0.048 | 0.036 | 0.084 | 0.036 | 0.084 |
| $O_2$ | -0.461 | -0.130 | -0.591 | -0.130 | -0.591 |
| Amino N | 0.199 | -0.460 | -0.260 | -0.460 | -0.260 |
| Amino H | | 0.191 | 0.191 | 0.191 | 0.191 |
| $H_5$ | | 0.033 | 0.033 | 0.033 | 0.033 |
| $H_6$ | | 0.037 | 0.037 | 0.037 | 0.037 |

TABLE 12 PART 3

CHARGE DISTRIBUTION

RIBOSE*

| Atom | Adenosine | Uridine | Guanosine | Cytidine | Pseudouridine |
|---|---|---|---|---|---|
| $C_1'$ | 0.127(0.116) | 0.127(0.116) | 0.127(0.116) | 0.127(0.116) | 0.078 |
| $H_1'$ | 0.054(0.053) | 0.054(0.053) | 0.054(0.053) | 0.054(0.053) | 0.049 |
| $O_1'$ | -0.271(-0.273) | -0.271(-0.273) | -0.271(-0.273) | -0.271(-0.273) | -0.279 |
| $C_2'$ | 0.107(-0.047) | 0.108(-0.047) | 0.108(-0.047) | 0.108(-0.047) | 0.102 |
| $H_2'$ | 0.052(0.041) | 0.052(0.041) | 0.052(0.041) | 0.052(0.041) | 0.051 |
| $O_2'$ | -0.458 | -0.458 | -0.458 | -0.458 | -0.459 |
| 2' Hydroxyl H | 0.301 | 0.302 | 0.302 | 0.302 | 0.301 |
| $C_3'$ | 0.104(0.093) | 0.104(0.093) | 0.104(0.093) | 0.104(0.093) | 0.103 |
| $H_3'$ | 0.052(0.050) | 0.052(0.050) | 0.052(0.050) | 0.052(0.050) | 0.052 |
| $O_3'$ | -0.458(-0.460) | -0.458(-0.460) | -0.458(-0.460) | -0.458(-0.460) | -0.458 |
| 3' Hydroxyl H | 0.301(0.301) | 0.301(0.301) | 0.301(0.301) | 0.301(0.301) | 0.301 |
| $C_4'$ | 0.094(0.093) | 0.904(0.093) | 0.904(0.093) | 0.904(0.093) | 0.094 |
| $H_4'$ | 0.051(0.050) | 0.051(0.050) | 0.051(0.050) | 0.051(0.050) | 0.050 |
| $C_5'$ | 0.045(0.045) | 0.045(0.045) | 0.045(0.045) | 0.045(0.045) | 0.045 |
| $H_5'$ | 0.053(0.053) | 0.053(0.053) | 0.053(0.053) | 0.053(0.053) | 0.053 |
| $O_5'$ | -0.457(-0.457) | -0.457(-0.457) | -0.457(-0.457) | -0.457(-0.457) | -0.457 |
| 5' Hydroxyl H | 0.302(0.302) | 0.302(0.302) | 0.302(0.302) | 0.302(0.302) | 0.302 |

* Numbers in parentheses are charges of the analogous atom in the deoxynucleoside. No calculations were done for deoxypseudouridine.

TABLE 12 PART 4

CHARGE DISTRIBUTIONS

| Atom | $\pi$ Charge | $\sigma$ Charge (Ribonucleoside) | Total Charge |
|---|---|---|---|
| | Pseudouridine | | |
| $N_1$ | 0.270 | -0.357 | -0.087 |
| $C_2$ | 0.194 | 0.254 | 0.448 |
| $N_3$ | 0.260 | -0.256 | 0.004 |
| $C_4$ | 0.185 | 0.023 | 0.208 |
| $C_5$ | -0.089 | 0.090 | 0.001 |
| $C_6$ | 0.015 | 0.049 | 0.064 |
| $H_1$ | | 0.200 | 0.200 |
| $O_2$ | -0.451 | -0.128 | -0.580 |
| $H_3$ | | 0.188 | 0.188 |
| $O_4$ | -0.384 | -0.084 | -0.468 |
| $H_6$ | | 0.039 | 0.039 |

TABLE 12 PART 5

CHARGE DISTRIBUTIONS

PHOSPHATE GROUP[38]

| Atom | Charge |
|---|---|
| P | 0.164 |
| O (P-O) | -0.552 |
| O (ester) | -0.350 |
| $C_3'$ | 0.164 less than the appropriate nucleoside $C_3'$ |
| $C_5'$ | 0.164 less than the appropriate nucleoside $C_5'$ |

TABLE 13

CHARGE DISTRIBUTIONS

| Atom | This Work | Ref. 13, 134 | Atom | This Work | Ref. 13, 134 |
|---|---|---|---|---|---|
| | Adenine | | | Cytosine | |
| $N_1$ | -.392 | -.392 | $N_1$ | .116 | .025 |
| $C_2$ | .249 | .276 | $C_2$ | .448 | .502 |
| $N_3$ | -.401 | -.399 | $N_3$ | -.415 | -.392 |
| $C_4$ | .161 | .195 | $C_4$ | .258 | .305 |
| $C_5$ | .015 | .036 | $C_5$ | -.146 | -.106 |
| $C_6$ | .295 | .316 | $C_6$ | .084 | .101 |
| $N_7$ | -.456 | -.416 | $O$ | -.591 | -.586 |
| $C_8$ | .154 | .194 | Amino N | -.260 | -.429 |
| $N_9$ | .169 | .031 | Amino H | .191 | .223 |
| Amino N | -.369 | -.442 | $H_5$ | .032 | .033 |
| Amino H | .218 | .223 | $H_6$ | .037 | .038 |
| $H_2$ | .042 | .042 | | | |
| $H_8$ | .043 | .043 | | | |
| | Guanine | | | Uracil | Thymine |
| $N_1$ | -.058 | -.052 | $N_1$ | .078 | -.044 |
| $C_2$ | .332 | .360 | $C_2$ | .441 | .526 |
| $N_3$ | -.505 | -.449 | $N_3$ | .002 | -.198 |
| $C_4$ | .102 | .224 | $C_4$ | .210 | .451 |
| $C_5$ | .022 | .059 | $C_5$ | -.055 | .005 |
| $C_6$ | .412 | .477 | $C_6$ | .057 | .044 |
| $N_7$ | -.405 | .375 | $O_2$ | -.581 | -.553 |
| $C_8$ | .118 | .162 | $H_3$ | .188 | .138 |
| $N_9$ | .189 | .052 | $O_4$ | -.469 | -.533 |
| Amino N | -.378 | -.431 | $H_5$ | .037 | |
| Amino H | .219 | .225 | $H_6$ | .038 | .038 |
| O | -.563 | -.577 | | | |
| $H_3$ | .201 | ($H_1$) .093 | | | |
| $H_8$ | .043 | .043 | | | |

The polarizability of phosphorus was obtained from the refraction of $PH_3$[184], subtracting the contribution of the hydrogens. The polarizabilities used in our calculation are listed in Table 14.

TABLE 14

| Atom | Polarizability ($Å^3$) |
|---|---|
| H | .42 |
| C | .93 |
| N (primary) | .87 |
| N (secondary) | .93 |
| N (tertiary) | 1.03 |
| O (hydroxyl) | .59 |
| O (ether) | .64 |
| O (carbonyl) | .84 |
| P | 3.00 |

($0.29Å^3$ is added to each atom on a double bond.)

Van der Waals radii were taken from the review by Bondi[12]. These are the values which are used in the calculation of the conformations of polypeptides and oligopeptides (with the exception of hard core repulsion studies). These atomic radii are listed in Table 15.

TABLE 15

| Atom | Van der Waals Radius (Å) |
|---|---|
| H | 1.20 |
| C | 1.70 |
| N | 1.55 |
| O | 1.52 |
| P | 1.80 |

The usual expression used for the London dispersion interactions is:

$$E_{ab} = \left(\frac{I_a I_b}{I_a + I_b}\right)\left(\frac{\alpha_a \alpha_b}{r_{ab}^{\ 6}}\right)$$

where $E_{ab}$ is the energy of interaction, $I_a$ and $I_b$ are the ionization energies of atoms a and b, $\alpha_a$ and $\alpha_b$ are the polarizabilities of atoms a and b and $r_{ab}$ is the interatomic distance. Instead of having to introduce the atomic ionization potentials required by this expression we preferred to use the expression proposed by Pitzer:[7]

$$E_{ab} = \frac{3e\hbar\alpha_a\alpha_b}{2m\ r_{ab}^{\ 6}\left[\left(\frac{\alpha_a}{N_{eff_a}}\right)^{\frac{1}{2}} + \left(\frac{\alpha_b}{N_{eff_b}}\right)^{\frac{1}{2}}\right]}$$

where e is the electronic charge; $\hbar$ is Planck's constant divided by $2\pi$; m is the electronic mass and $N_{eff}$ is a parameter describing an atom which can be considered to be the "effective number of electrons". In this expression we can use one value of $N_{eff}$ for each atom, instead of a

series of ionization potentials for each type of atom (e.g. primary, secondary and tertiary nitrogens). The values of $N_{eff}$ were determined by fitting all of the experimental data in Pitzer's review[144] to a straight line. The best line by the least squares criterion going through to origin is:

$$N_{eff} = 1.146 \text{ (Atomic Number)}$$

These numbers are somewhat larger (up to 25%) than those used in calculations on polypeptides[19,170,171], which were based entirely on $N_{eff}$ reported for rare gases. We don't feel that the differences between these sets of $N_{eff}$ will affect our results very much because this number comes in as a square root, rather than being raised to a high power.

Since the choice of parameters for these calculations is to a large part arbitrary we should compare these values with others systematically chosen on the basis of other criteria. The only study with which we are familiar is a survey of intermolecular forces in terms of $r^{-6}$ attraction and an exponential repulsion in crystals of hydrocarbons[217]. (The survey was not able to determine these coefficients unambiguously. The four best sets of coefficients are reported, and listed in Table 16.)

TABLE 16

COEFFICIENTS OF $R^{-6}$ ATTRACTIONS (in kcal/$Å^6$)

| C-C | C-H | H-H | FIT* |
|---|---|---|---|
| -602 | -86 | -71 | 131 |
| -440 | -160 | -58 | 301 |
| -437 | -168 | -50 | 302 |
| -535 | -139 | -36 | 186 |
| -645 | -182 | -53 | (parameters used in these calculations) |

* The standard deviations of the experimental and calculated crystal parameters.

In a calculation as complicated as this it is important that scrupulous checks be made for incorrect data or "minor" arithmetic errors which might not be obvious, but could still invalidate the results. The atomic coordinates, polarizabilities and "identification" numbers (used to assign the appropriate Van der Waals radii, etc.) were checked frequently against the original data. This was very important, as data decks were duplicated frequently. The charges were checked against the original data and also added to see if they totaled 0 in the case of nucleosides or -1 in the case of CpC. The routine which changes atomic coordinates in accordance with rotations about single bonds was tested by rotating by 120° 3 times, or 20° 18 times, or by 10° 36 times and being sure that the atoms returned to their original coordinates. Also, starting at the standard orientation, rotations were made around bonds to reproduce the conformations found in crystal structures. These were checked against the atomic

coordinates of the structures which were calculated independently. Hand calculations of single interatomic interactions were made and used to verify results of the energy calculations. In the cases where derivatives were calculated analytically, they were checked against derivatives calculated by computing the difference in energy between two conformations differing only on a small change in one torsional angle (in the case of a first derivative), etc.

Calculations were performed on either the CDC 6600's at the Lawrence Radiation Laboratory, Berkeley, or the CDC 6400 at the University of California Computing Center. All programs used were written in Chippewa FORTRAN. Listings and descriptions of the important programs and subroutines used in these studies are given in Appendix B.

## RESULTS AND DISCUSSION

### NUCLEOSIDES

Our first calculations were the variation of energy with changing $\phi_{CN}$. This is a good place to start because a number of checks for our calculations already exist. The crystal structures of a number of nucleosides, nucleotides and other suitable compounds have already been determined[23,57,58,73,76,77,82,83,99,172,173,181,182,193,212]. In addition, we expect the ORD of nucleosides in the region of strong base absorption (~260 mμ) to be very sensitive to the orientation of the base with respect to the sugar. The ORD originates in the interaction of the base with the asymmetric ribose. Calculation of the ORD

of nucleosides as a function of $\phi_{CN}$ are presently being undertaken[84]. In addition, there are experimental studies of the ORD of nucleosides and suitable model compounds (in which the base is locked into a conformation by covalent linkages) which can be used to make meaningful checks of our calculations[53,54,98,128,201,219].

The first nucleosides for which calculations were made were cytidine and uridine. In order to consider the possibility that the furanose ring was not in the 2' endo conformation, calculations for 3' endo conformations were carried out as well. We also calculated the potential energy for rotation about the glycosidic bond for the 2' endo and 3' endo deoxyribose analogues. The results of these calculations are presented in Figures 29-32.

These results are clear cut. The pyrimidine nucleosides are much more stable in the anti conformation, with the syn conformation approximately 5 kcal per mole higher. In these calculations the orientation of the 2' OH was that found in the crystal structure from which the atomic coordinates were taken. The possibility existed that the most stable conformation of the isolated nucleoside would be with the base syn and the 2' OH oriented in a direction quite different from that found in crystal structures. To eliminate this possibility we performed the following calculation: Cytidine and uridine were oriented in each of four conformations, with bases syn or anti and the 2' OH aligned approximately as it is found in the crystal or with a 180° rotation about the $C_2'$-$O_2'$ bond. Then rotations were permitted about the glycosidic and $C_2'$-$O_2'$ bonds which

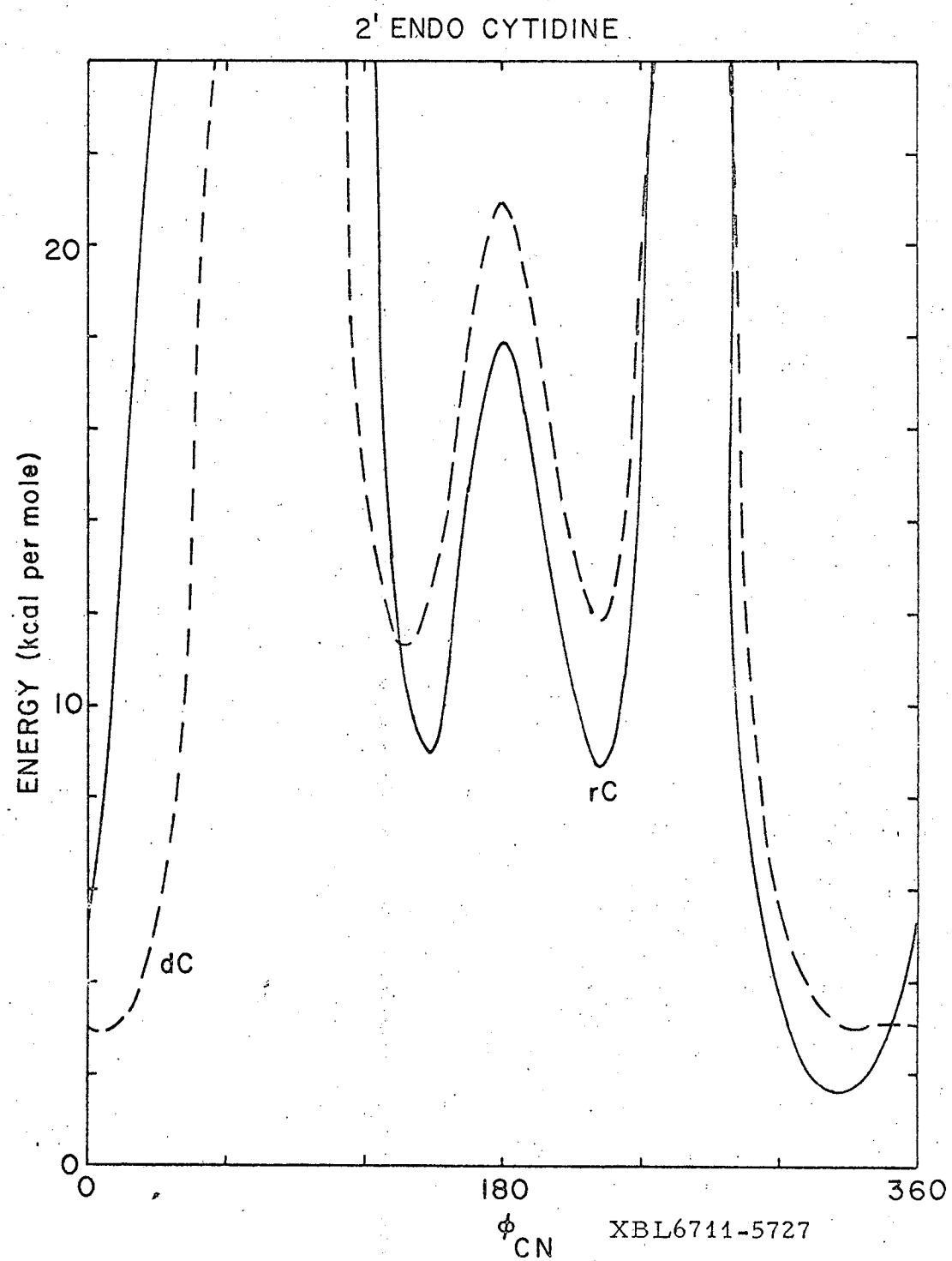



Figure 29 Calculated energy as a function of $\phi_{CN}$ for 2' endo cytidine and deoxycytidine.

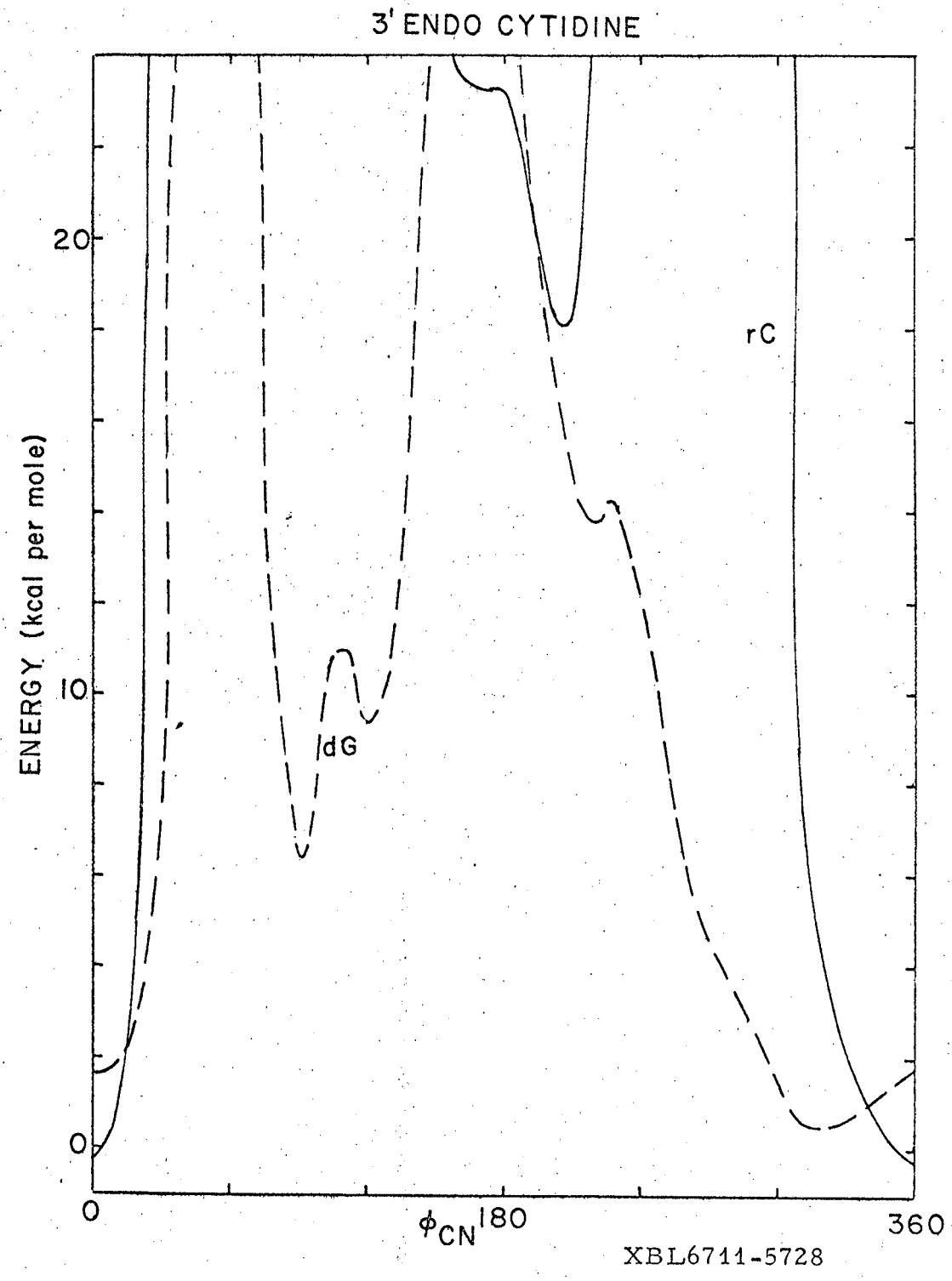



Figure 30 Calculated energy as a function of $\phi_{CN}$ for 3' endo cytidine and deoxycytidine.

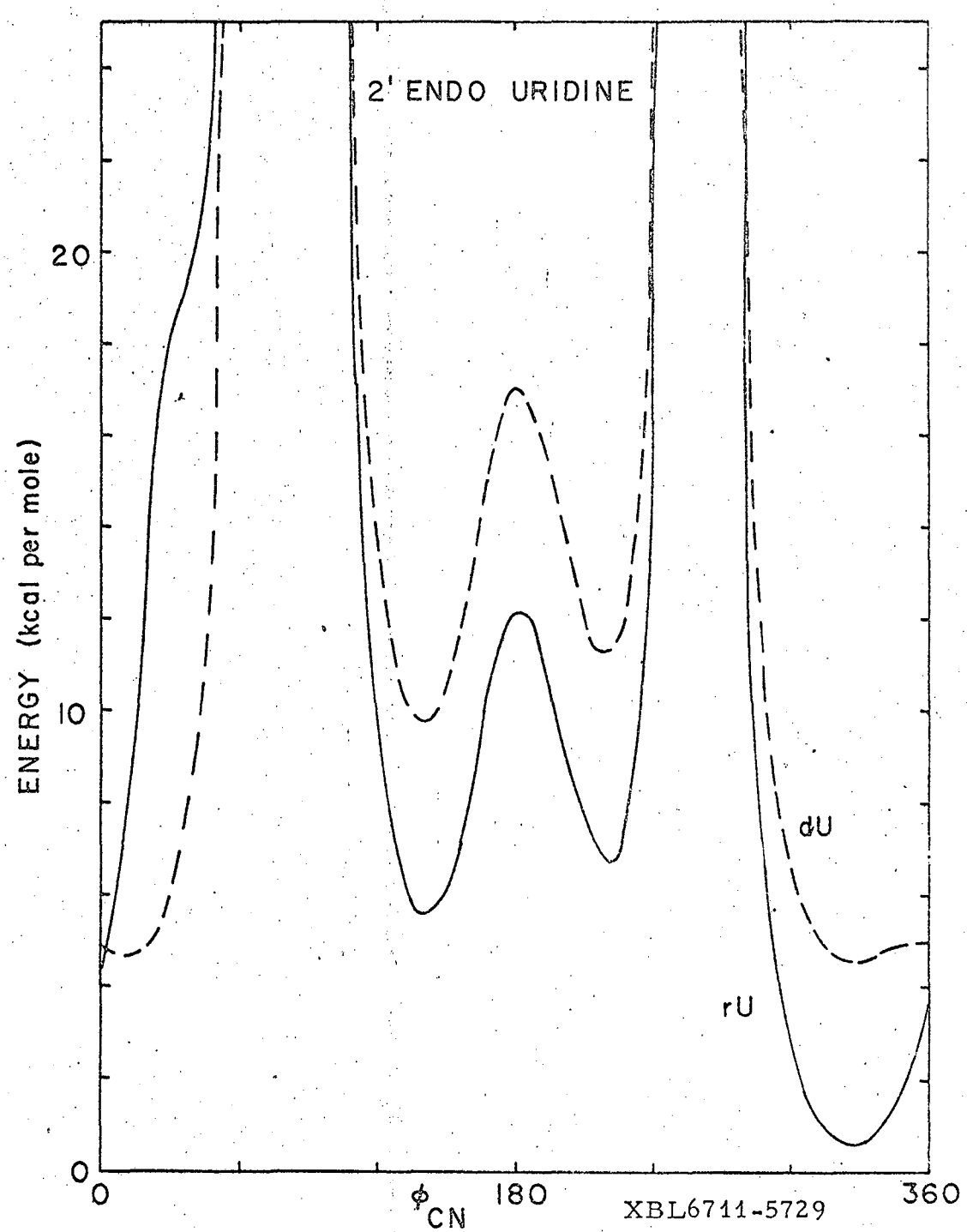



Figure 31 Calculated energy as a function of $\phi_{CN}$ for 2' endo uridine and deoxyuridine.

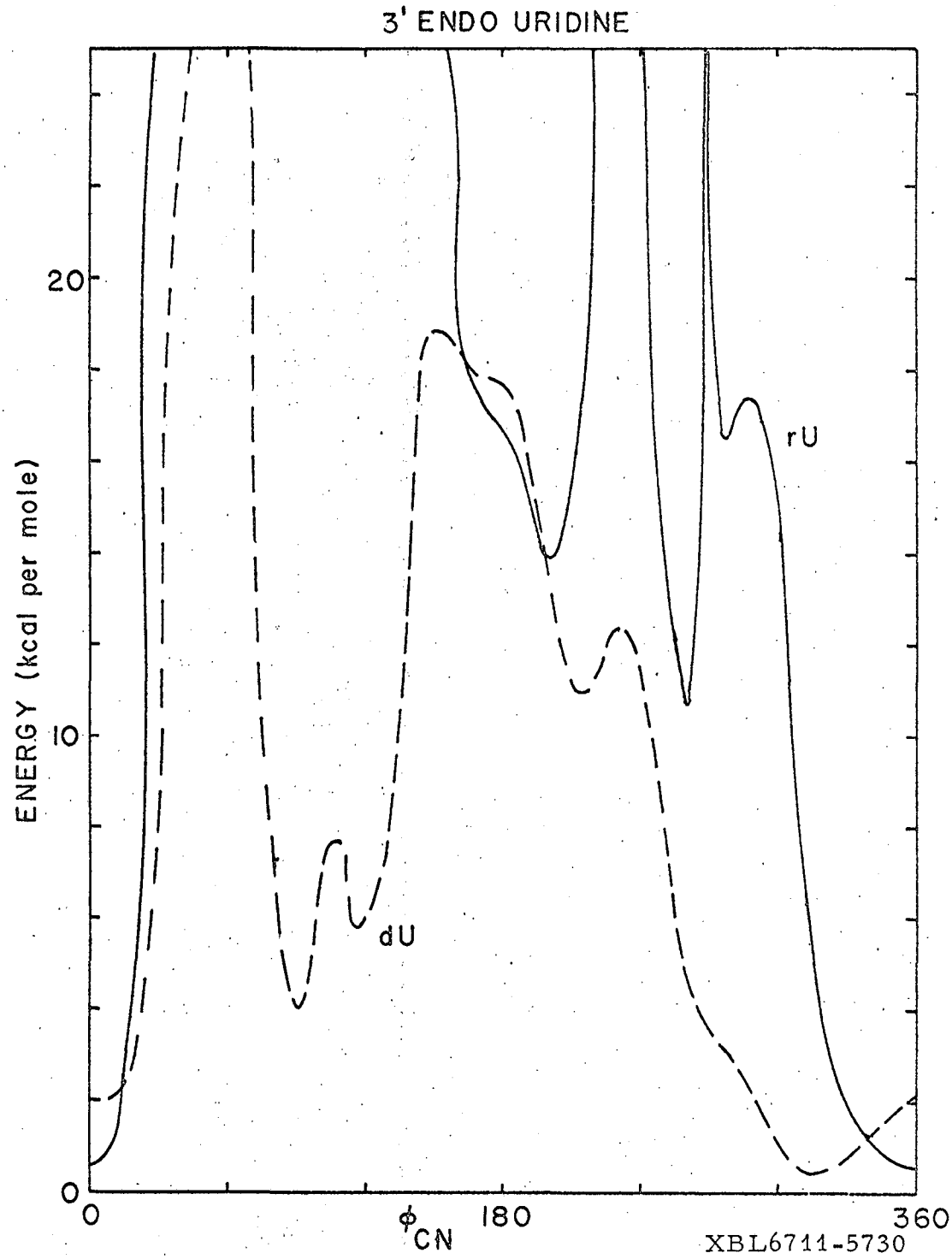



Figure 32 Calculated energy as a function of $\phi_{CN}$ for 3' endo uridine and deoxyuridine.

enable the compound to minimize its energy and seek a potential minimum. (The program, FINDMIN, was used.) These calculations indicate that the anti forms of cytidine and uridine are energetically more stable by approximately 5 kcal per mole.

The origin of this energy difference is reasonably straight forward. In the anti conformation, the small, almost uncharged, $H_6$ is close to the large and negatively charged $O_1'$. The repulsion between these two atoms is relatively small. In the syn conformation, the large, negatively charged $O_2$ is close to $O_1'$. There is appreciable repulsion between these two atoms, both steric and coulombic. For this reason, the puckering of the furanose ring (so long as $O_1'$ is not puckered) or the substitution of deoxyribose for ribose does not change this difference very much.

We have also done similar calculations for 2' endo pseudouridine. (To the best of our knowledge no other studies of its conformation have been undertaken.) We thought that the longer glycosidic bond (1.54Å for the C-C bond rather than 1.47Å for the C-N bond) might substantially eliminate the unfavorable close contacts which tend to destabilize the syn conformation relative to the anti. Apparently this reduction is not significant. Again, the anti conformation is more stable than the syn form by approximately 4.5 kcal per mole. (See Figure 33.)

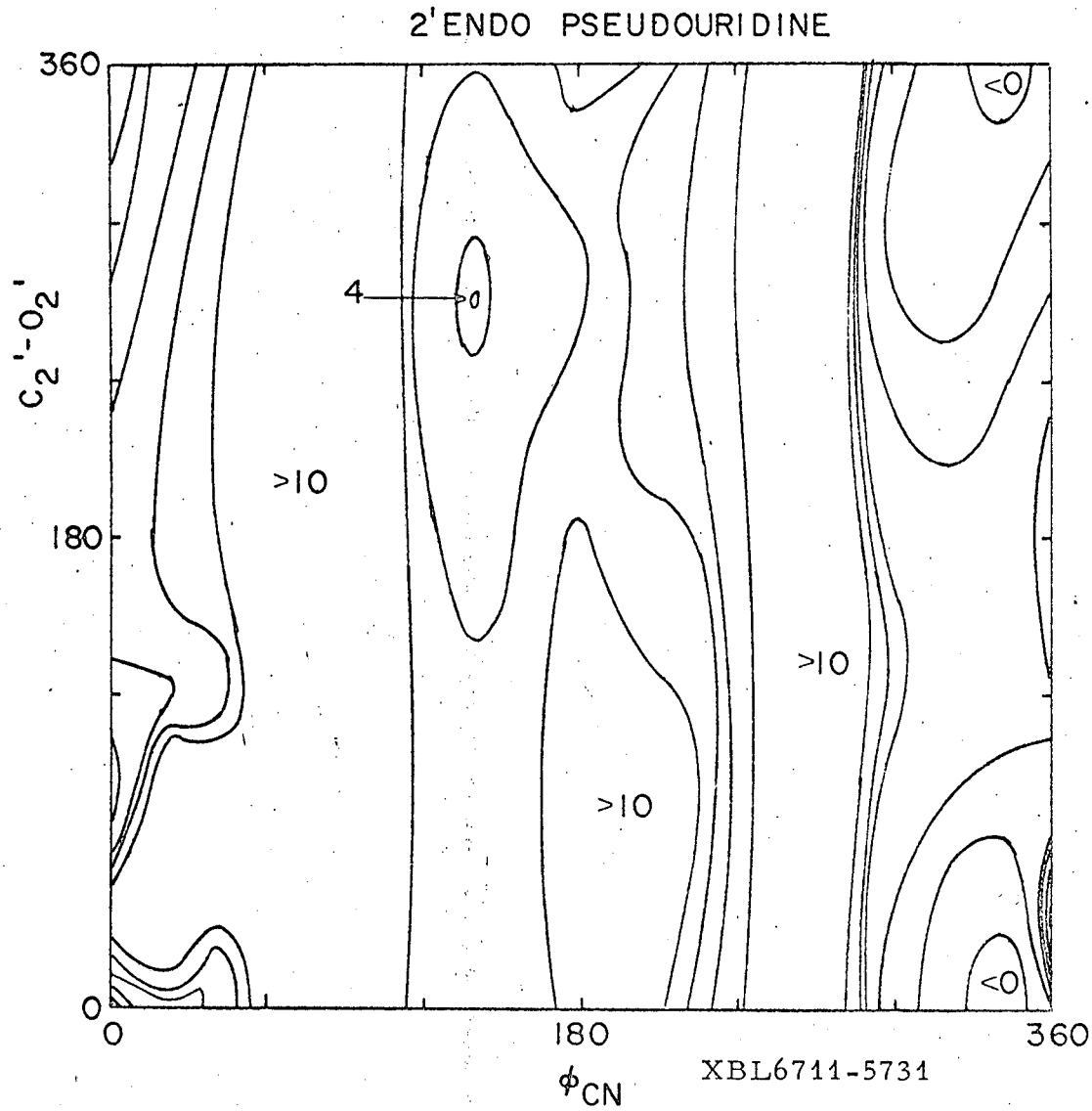



Figure 33 Calculated energy as a function of $\phi_{CN}$ and the $C_2'-O_2'$ torsional angle for 2' endo pseudouridine. Contours are at every 2 kcal per mole.

The major results of our calculations can be summarized quite simply. The anti forms of pyrimidine nucleosides are considerably more stable than the syn forms, with both 2' endo and 3' endo ribose and deoxyribose. We expect this will apply to other pyrimidine nucleosides as well. In the case of pseudouridine this has implications about its role in determining the three dimensional structure of transfer RNAs. It is possible that pseudouridine may participate in hydrogen bonds in which uridine can not. $H_1$ can participate in hydrogen bonds as well as $H_3$, $O_2$, and $O_4$ (which are also present in uridine). It is possible that pseudouridine can hydrogen bond with two different bases, thereby forming the nucleus of a triple-strand structure, something uridine can not do. $H_3$ and $O_2$ can participate in hydrogen bonding schemes similar to those of uridine. $H_1$ and $O_2$ can also form a second set of hydrogen bonds.

There is one result of these calculations which must be interpreted carefully. The energy barrier for the anti to syn exchange is high, about 30-50 kcal per mole, which corresponds to approximately half of bond formation energies. There will be appreciable distortions of a "rigid" group with that large a force put on it. These distortions will tend to minimize the total potential energy of the system. The barrier heights we calculate are upper limits (assuming our calculations have been executed correctly). Calculation of the actual barrier heights requires knowledge of the potential energy for large

bending and stretching movements which is not available. For this reason we can not say at this time how fast the exchange is between the syn and anti forms of pyrimidine nucleosides. The possibility does exist that such an interchange may be very slow. If this is the case then syn conformations in pyrimidine nucleosides will not be significant even as biological intermediates because their formation and return to the more stable anti conformation are too slow.

A number of experimental studies have also found that pyrimidine nucleosides are in the anti conformation. Of the large number of pyrimidine nucleosides and nucleotides for which molecular conformations have been determined by X-ray diffraction techniques, all have been in the anti conformation, with values of $\phi_{CN}$ which we calculate to have low energies. These results are summarized in Table 17.

Measurements of solution properties also agree with our results. The ORD of pyrimidine nucleosides and pyrimidine nucleosides in which covalent ribose-base bridges lock the base into the anti confromation (e.g. $O_2$-2' cyclouridine) are substantially the same[54,201]. This implies that their conformations are the same (i.e. anti).

In addition, a model-building consideration of nucleosides (in which the basic criterion for stability is maximization of interatomic distances smaller than the usual Van der Waals interatomic distances) also predicts that the anti conformation should be considerably more stable than the syn conformation for pyrimidine nucleosides[75].

TABLE 17

| Base | Crystal | Furanose Puckering | $\phi_{CN}$ | Reference |
|---|---|---|---|---|
| C | 3' Cytidylic Acid | 2' endo | -42.1 | 182 |
| C | 3' Cytidylic Acid | 2' endo | -39.9 | 23 |
| C | Cytidine | 3' endo | -18 | 58 |
| 5BrC | 5 Bromodeoxycytidine + Deoxyguanosine | 2' endo | -61 | 77 |
| U | 5' Uridine phosphate, Ba salt | 2' endo | -43 | 173 |
| U | Adenylyl (2'-5') Uridine | 3' endo | - 5 | 172 |
| 5BrU | Adenosine + 5 Bromouridine | 3' endo | -20 | 76 |
| 5BrU | 5 Bromouridine | 2' endo | -43 | 83 |
| 5BrU | 5 Bromodeoxyuridine | 2' endo | -56 | 83 |
| 5FU | 5 Fluorodeoxyuridine | 2' endo | -60 | 73 |
| T | Ca Thymidylate | 3' endo | -48 | 193 |
| | CALCULATED | | | |
| | Cytidine | 2' endo | -33 | |
| | Deoxycytidine | 2' endo | -30 to 5 | |
| | Cytidine | 3' endo | 0 | |
| | Deoxycytidine | 3' endo | -42 | |
| | Uridine | 2' endo | -32 | |
| | Deoxyuridine | 2' endo | -30 | |
| | Uridine | 3' endo | - 3 | |
| | Deoxyuridine | 3' endo | -45 | |
| | Pseudouridine | 2' endo | -20 | |

The situation is quite different in the case of purine nucleosides. Figures 34-37 show the potential for rotation about the glycosidic bond for guanosine and adenosine. The 2' OH (in the case of ribose) is in the orientation that is found in the crystal strucutre from which the atomic coordinates of the ribose were taken. There is only a small difference in energy between the syn and anti conformations. Because this difference is so small it is necessary to be sure that the torsional angle for the $C_2'$-$O_2'$ bond is in its potential minimum as an incorrect orientation of the $C_2'$-$O_2'$ bond might shift the energies enough to exchange the relative stabilities of the two conformations. The results of calculations of the two dimensional energy surface for rotations about both the glycosidic and $C_2'$-$O_2'$ bonds for 2' endo adenosine and 2' endo guanosine are presented as energy contour diagrams. (Figures 38 and 39).

In the cases of both 2' endo adenosine and 2' endo guanosine the syn conformation is slightly more favorable than the anti form. The energy differences for guanosine and adenosine are approximately 1 kcal per mole and less than 1 kcal per mole, respectively. The potential energy barriers for the syn-anti interchange are low, about 5 kcal per mole for adenosine and 6 kcal per mole for guanosine. Therefore interchange between the two conformations should be very rapid. The energetically unfavorable conformation may still be biologically important as an intermediate in

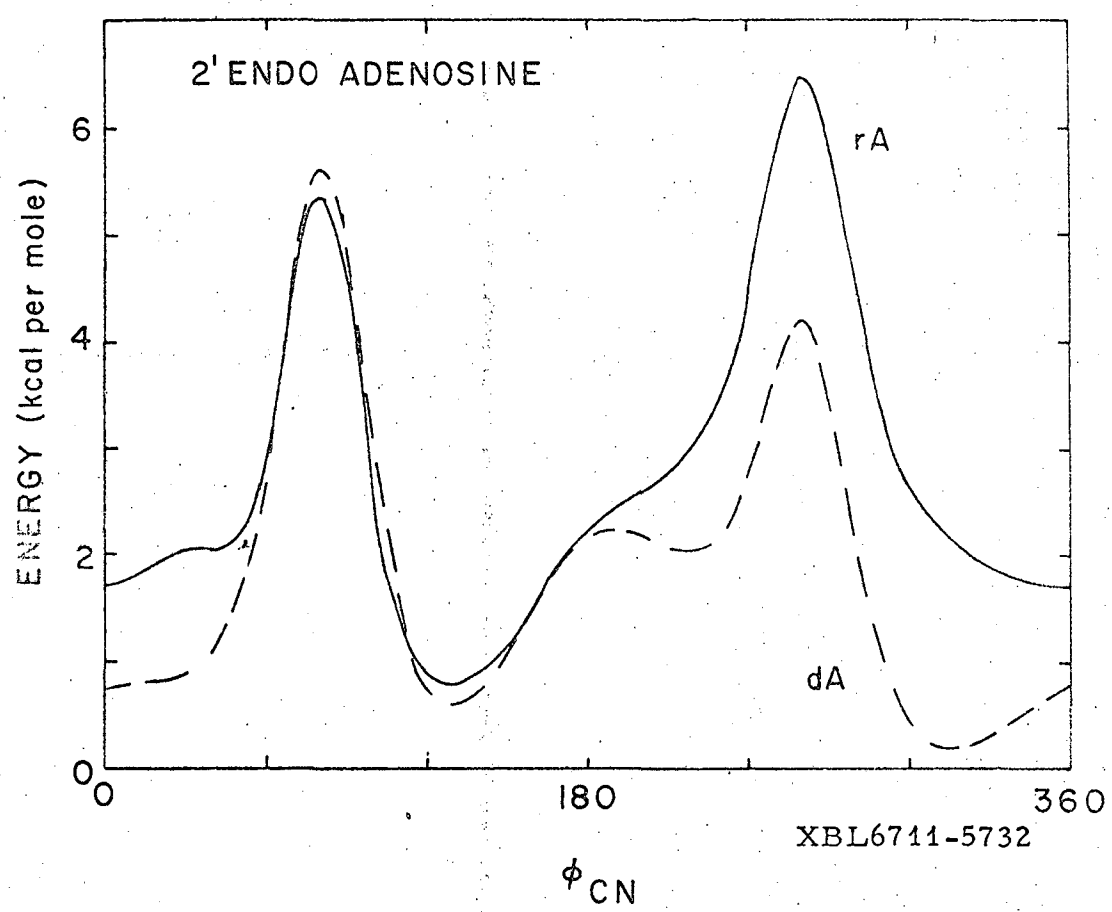



Figure 34 Calculated energy as a function of $\phi_{CN}$ for 2' endo adenosine and deoxyadenosine.

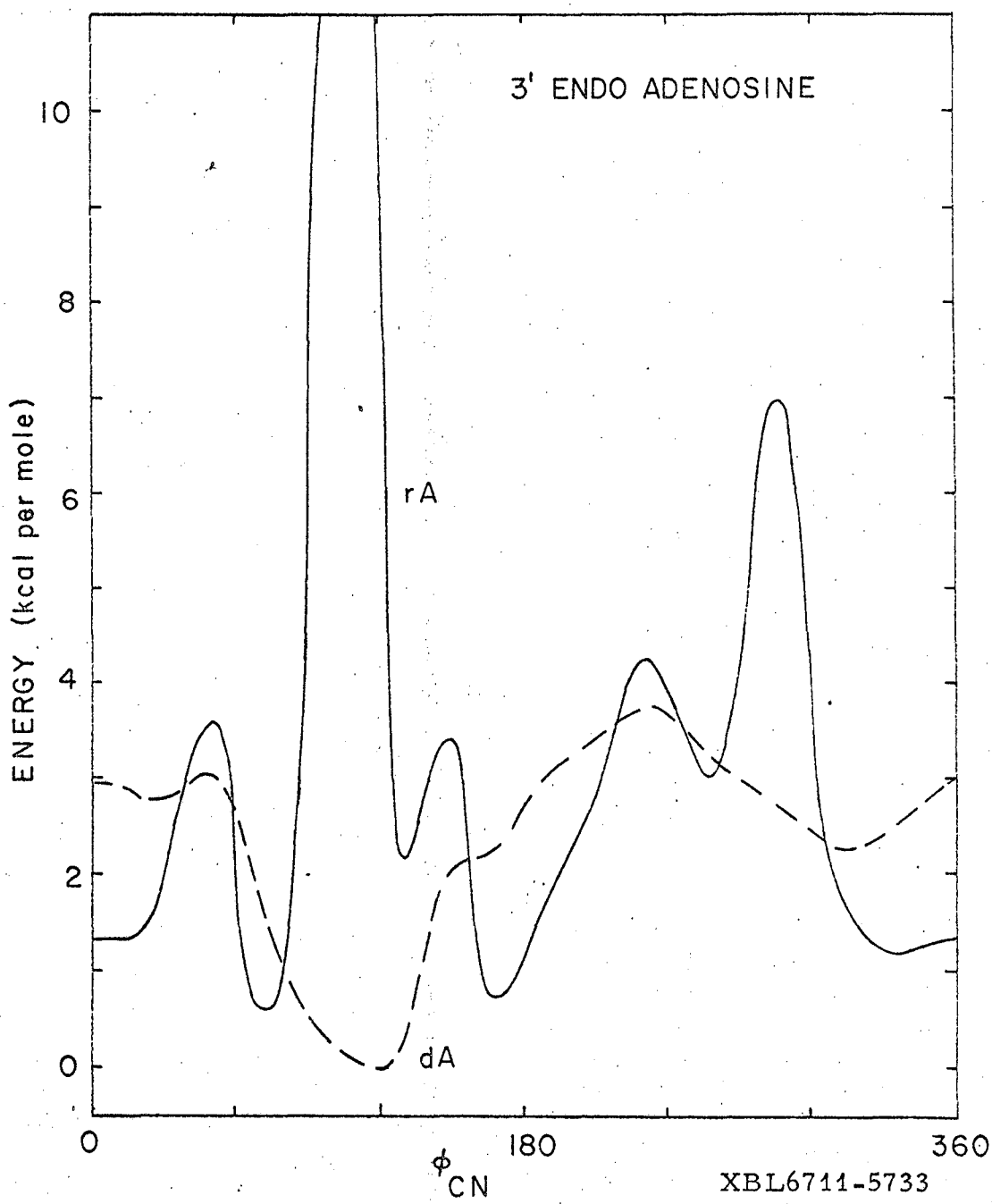



Figure 35 Calculated energy as a function of $\phi_{CN}$ for 3' endo adenosine and deoxyadenosine.

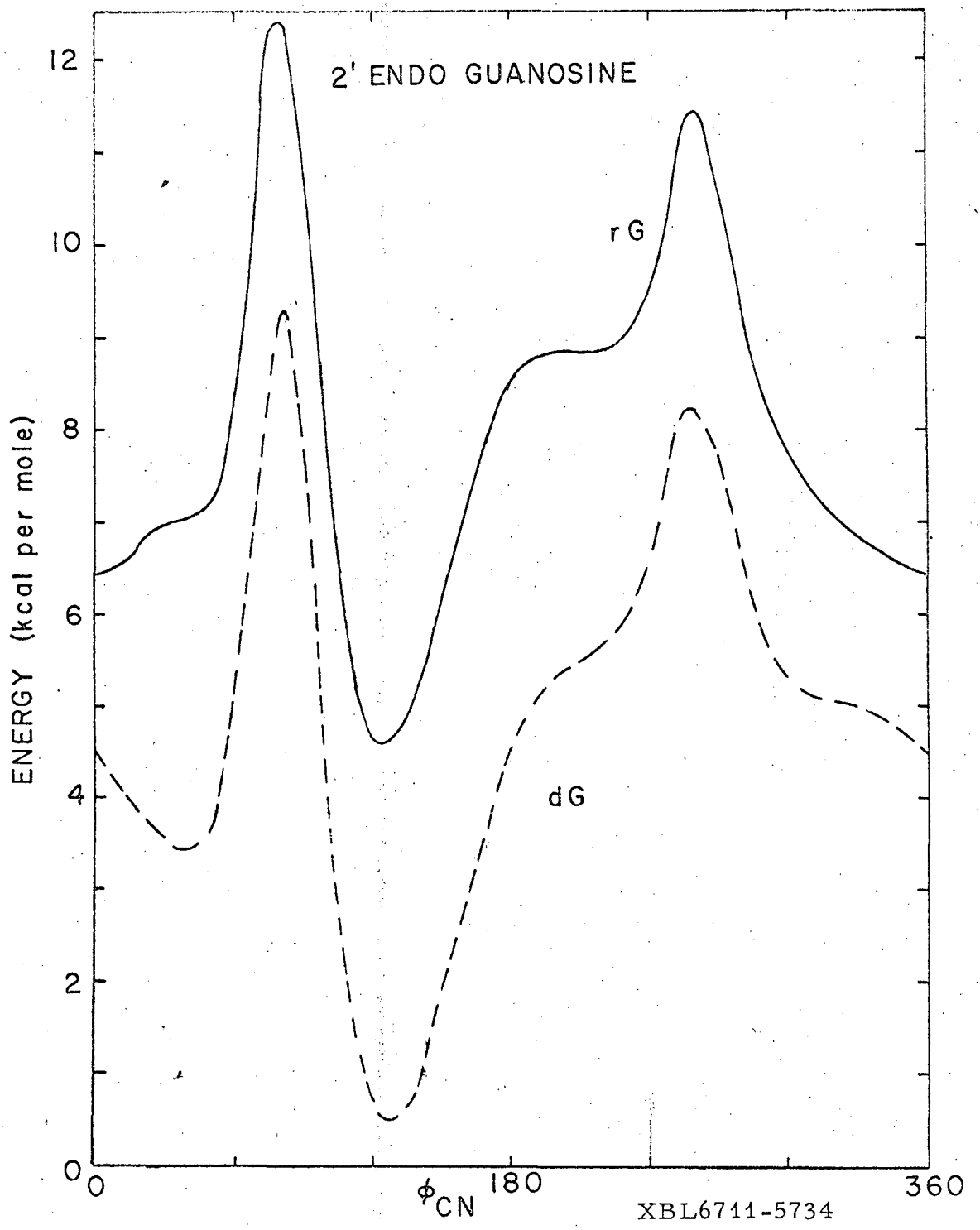



Figure 36 Calculated energy as a function of $\phi_{CN}$ for 2' endo guanosine and deoxyguanosine.

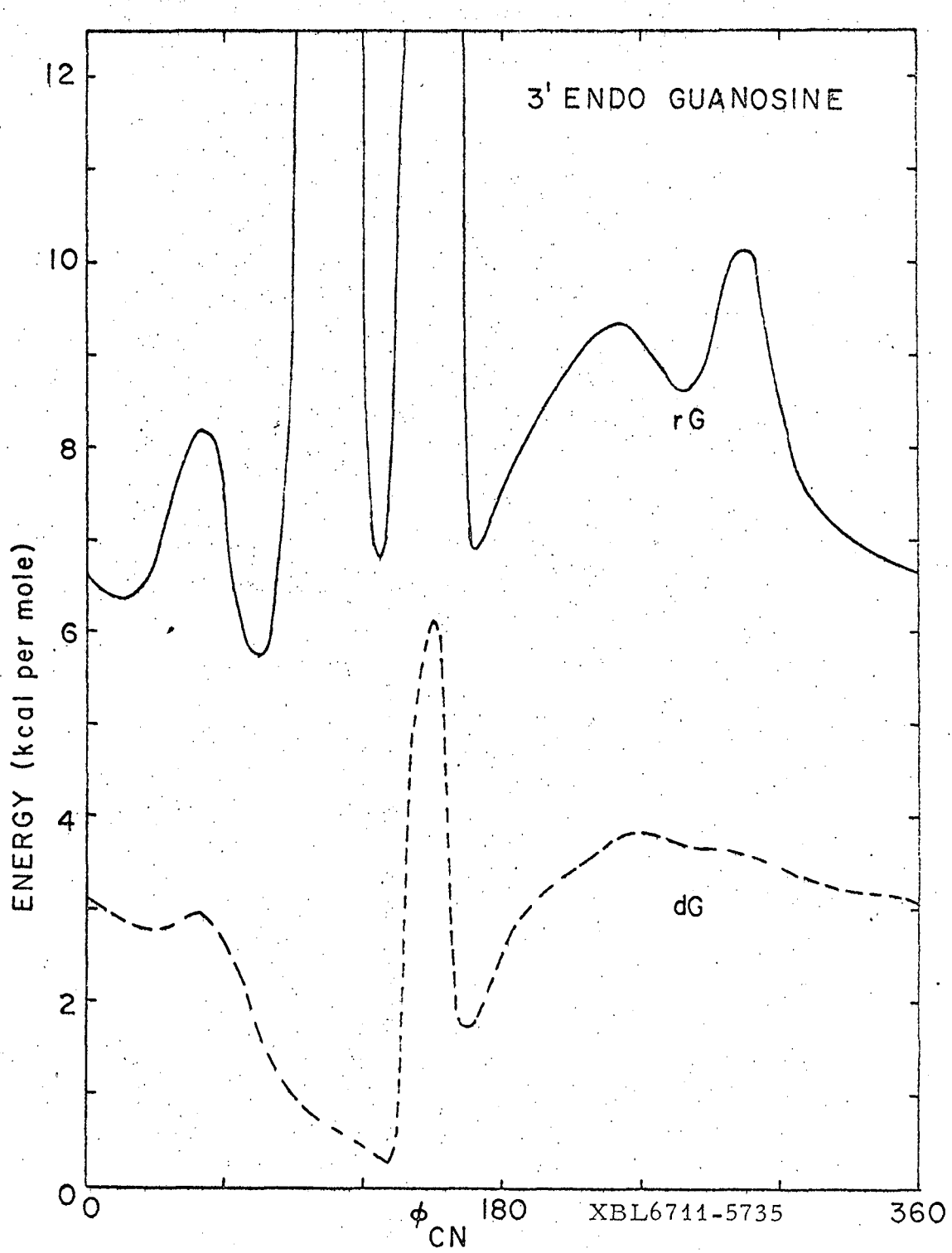



Figure 37 Calculated energy as a function of $\phi_{CN}$ for 3' endo guanosine and deoxyguanosine.

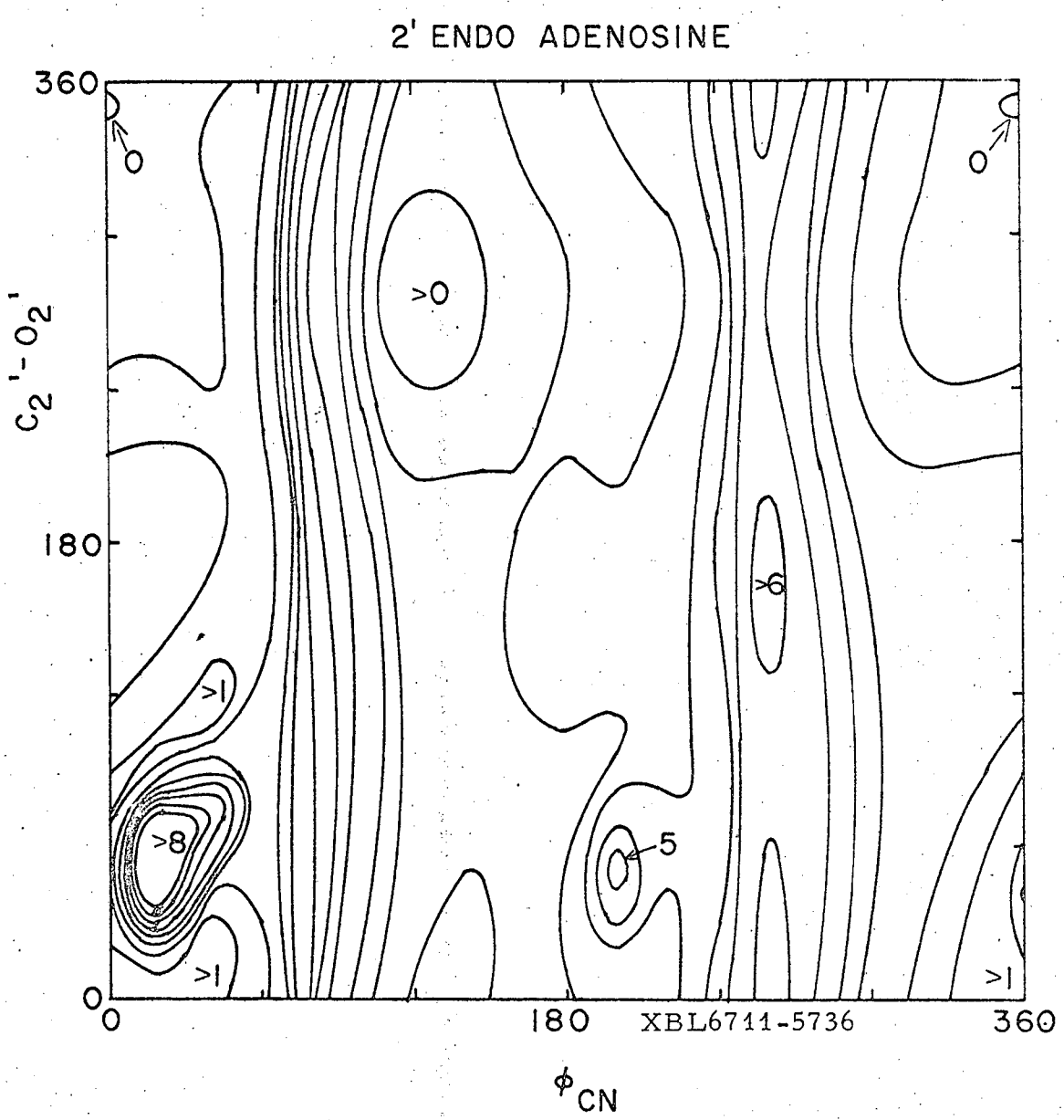



Figure 38 Calculated energy as a function of $\phi_{CN}$ and the $C_2'-O_2'$ torsional angle for 2' endo adenosine. Contours are at every 1 kcal per mole.

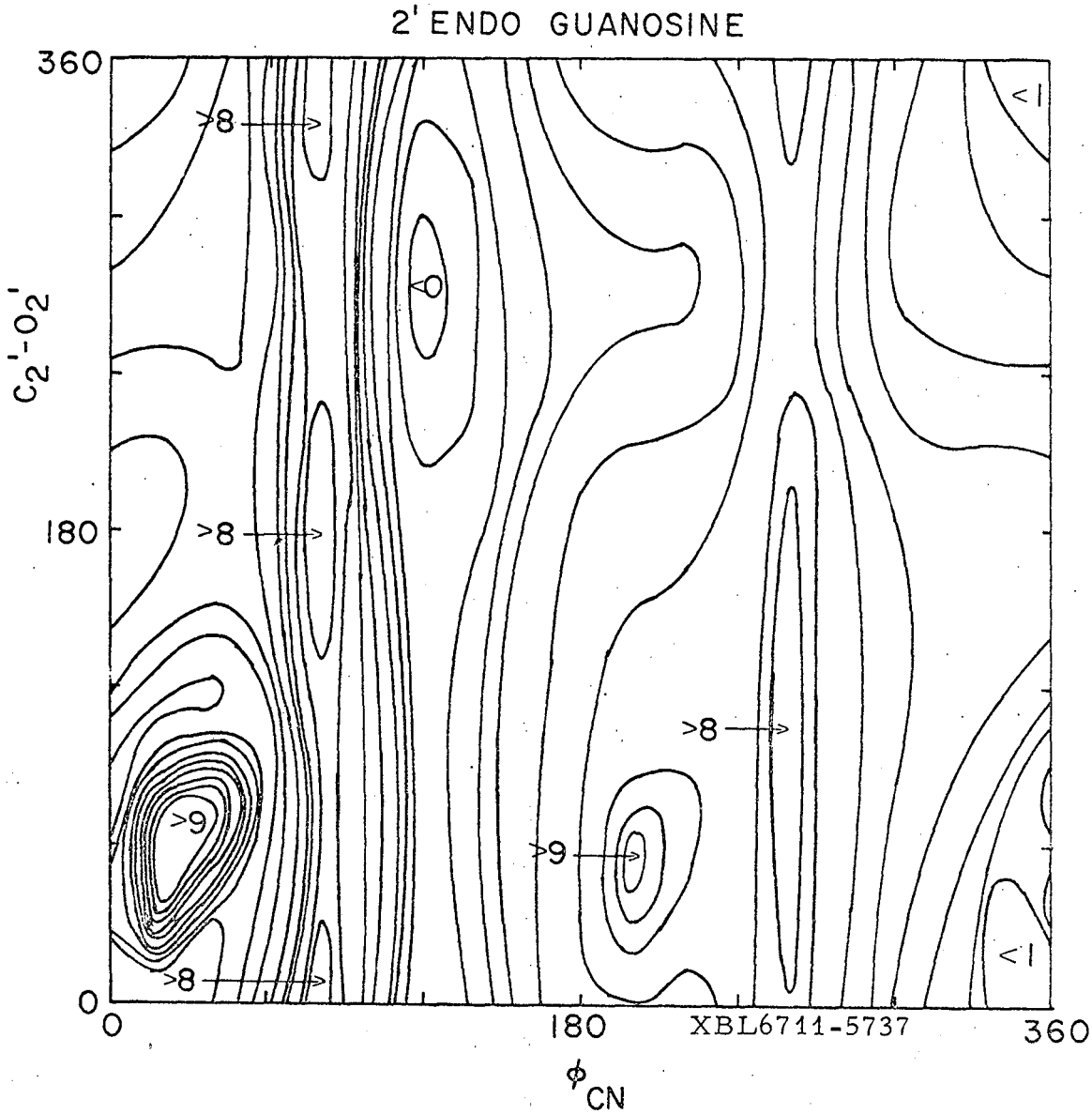



Figure 39 Calculated energy as a function of $\phi_{CN}$ and the $C_2'-O_2'$ torsional angle for 2' endo guanosine. Contours are at every 1 kcal per mole.

a reaction, as its formation and return to the favorable conformation will not appreciably slow a reaction (as may be the case with pyrimidines).

There are important differences between adenosine and guanosine and their deoxyribose analogues. The syn forms of deoxyadenosine and deoxyguanosine are not always more stable than the anti forms. The differences between the syn and anti forms (which were about 1 kcal per mole for 2' endo adenosine and guanosine) now range from the anti form's being more stable by 1 kcal per mole (2' endo deoxyadenosine) to the syn form's being more stable by 5 kcal per mole (3' endo deoxyadenosine). This may have an important effect on the structure of ribo and deoxyribo-polynucleotides. Let us assume that the orientation of a base in a single-strand polynucleotide (e.g. G in a single-strand DNA or RNA) is the same as in the nucleoside (i.e. syn). (Although this is by no means a certainty, it is more likely than not.) In order to form Watson-Crick complementary base-pairs (e.g. double-strand RNA or DNA) the bases must flip over to the anti form. (We already know that pyrimidines are in the anti-conformation. In order to maintain the symmetry of the double-strand helix purines must assume this conformation also.) Although energy is lost in flipping the base from syn to anti, more than enough energy is gained in double-strand formation to make up for it. But it takes more energy to flip guanine from syn to anti in dG than it does in rG.

Therefore if all other effects are equal the same RNA complex will be more stable than the DNA complex relative to the single-strand species with the same conformation. It is found that complementary double-strand polyribonucleotide complexes are more stable (have higher $T_m$) than their deoxyribose analogues[31,32]. Of course, everything else is not the same in both cases. Single-strand polyribonucleotides probably have slightly different structures than do their deoxyribo analogues[197,206]. Also complementary double-strand RNAs and DNAs have slightly different structures (as indicated by X-ray diffraction studies)[4]. Obviously the energetic differences between syn and anti conformations in ribonucleosides and deoxyribonucleosides are not the only important effects. In any case, they should be considered.

X-ray crystallographic studies of adenosine and adenylic acid indicate that the anti conformation is more stable[77,99,172,181,212]. We predict that the syn conformation should be slightly more favored than the anti. Intermolecular hydrogen bonding or crystal packing forces could easily make the anti conformation more stable.

Only one structure of a nucleoside or nucleotide containing guanine has been published up to this time[77]. $\phi_{CN}$ for doexyguanosine (hydrogen bonded to 5 bromodeoxycytidine) is in the syn region.

The results of crystal structure studies of purine nucleosides and nucleotides are summarized in Table 18.

TABLE 18

| Base | Crystal | Furanose Puckering | $\phi_{CN}$ | Reference |
|---|---|---|---|---|
| A | 5' Adenylic acid | 3' endo | -18 | 99 |
| A | Deoxyadenosine | 3' exo | - 3 | 212 |
| A | 3' Adenylic Acid | 3' endo | - 3.9 | 181 |
| A | Adenylyl (2'-5') Uridine | 2' endo | -55 | 172 |
| A | Adenosine + 5 Bromouridine | 3' endo | -10 | 76 |
| G | 5 Bromodeoxycytidine + Deoxyguanosine | 2' endo | +138 | 77 |
| | CALCULATED | | | |
| | Adenosine | 2' endo | +128 | |
| | Deoxyadenosine | 2' endo | -45 | |
| | Adenosine | 3' endo | +73 | |
| | Deoxyadenosine | 3' endo | +118 | |
| | Guanosine | 2' endo | +120 | |
| | Deoxyguanosine | 2' endo | +128 | |
| | Guanosine | 3' endo | +75 | |
| | Deoxyguanosine | 3' endo | +130 | |

Both syn and anti orientations are found indicating that there probably isn't a large energy difference between the two. That agrees with our calculations.

Some ORD studies similar to those on pyrimidine ribosides seem to indicate that the anti orientation is more favorable[128]. However, the evidence is not nearly as clear as it was in the case of pyrimidine ribosides. Purine riboside optical rotations are so small that perturbations on the chromophore by the covalent bridges which lock the base rigidly into the syn or anti conformation, may be large enough to change the sign of the rotational strength associated with the base absorption. Much of this ambiguity may also result from the lability of these compounds[98]. Arguments based on the ORD of adenosine with various substitutions on the 5' carbon (which are bulky enough to force the base into the anti conformation) indicate that the orientation of the base in adenosine is syn[98]. NMR studies on ApA indicate that both bases are in the anti conformation, but that the syn conformation is approximately kT higher in energy[34].

The ambiguity of these studies still leaves unanswered the question of the orientation about the glycosidic bond in purine ribosides. Our results seem reasonable; the syn and anti conformations of A and G are very close in energy. It will be interesting to see if a syn adenosine conformation or an anti guanosine conformation will be

found in future X-ray diffraction studies. Hopefully we will be able to test our results more critically when calculations of the ORD of nucleosides (as a function of $\phi_{CN}$) are available soon.

Structures of polynculeotides determined by fiber diffraction techniques also offer checks for our results. The two proposed structures for ribosomal RNA fragments[62,178], the acid form of poly A[161] and the A form of DNA[63] all have values of $\phi_{CN}$ which we found to have low energies. $\phi_{CN}$ ranges from -14.4° to 0.9° for these structures. The B and C forms of DNA[104,122] seem to have anomalous values for $\phi_{CN}$, -86.1° and -73.3°, respectively, considerably more negative than those found in X-ray crystal studies of nucleosides or nucleotides. These values of $\phi_{CN}$ seem to correspond to regions of maximum rather than minimum energy. We have no explanation for this discrepancy presently.

INTERNUCLEOSIDE BONDS

Advancing to more complicated systems, we made some calculations of potentials for rotating around the bonds which connect furanose rings (i.e. $C_3'$-$O_3'$, $O_3'$-P, P-$O_5'$, $O_5'$-$C_5'$ and $C_5'$-$C_4'$). Only those atoms on either side of the bonds interacting in a way completely determined by the one or two torsional angles being treated were considered. For example, if rotation around the $C_3'$-$O_3'$ bond were being considered as the only degree of freedom, then $H_3'$ or $H_4'$ of the ribose could interact with the

phosphate phosphorus, but not the phosphate oxygens (as their positions also depend on the orientation of the $O_3'$-P bond). Since the bonds in the middle of the phosphodiester bridge have a few atoms close to them their potential barriers for rotations are low. However, bonds near the furanose rings (e.g. $C_3'$-$O_3'$) have many atoms near them and may have orientations of high energy. Elimination of orientations of these bonds now because of their high energy, greatly reduces the large number of conformations we would have to try later when calculating the energies of dinucleoside phosphate conformations. For these reasons we have calculated two dimensional energy surfaces for the $C_3'$-$O_3'$, $O_3'$-P (both ribose and deoxyribose) and $C_4'$-$C_5'$, $C_5'$-$O_5'$ systems. In the former case the following atoms were considered to interact: $C_3'$, $O_2'$, 2' hydroxyl H (or $H'_2'$), $H_3'$, $H_4'$, $H_5'$, $H'_5'$, $O_5'$, 5' hydroxyl H, $O_3'$, P, phosphate O, phosphate O', $O_5'$. In the latter case the following atoms were considered to interact: $C_4'$, $H_4'$, $O_1'$, $H_3'$, $O_3'$, 3' hydroxyl H, $C_5'$, $H_5'$, $H'_5'$, $O_5'$, and P. It should be noted that not all atoms which have their positions defined (by the bond torsional angles being considered) are included. Only those which were thought to be able to interact strongly were included. For example, the distance between the phosphate P and $C_1'$ of the 5' linked nucleoside is determined by the orientation angles of the $C_4'$-$C_5'$ and $C_5'$-$O_5'$ bonds. However, it is highly unlikely that there will be a strong interaction between them. On the other hand, it is easy to see that there will be a strong inter-

action between the phosphate oxygens and the $O_2'$ of the 3' linked nucleoside in the $C_3'$-$O_3'$, $O_3'$-P system.

The results of these calculations are shown in Figures 40-42. Figure 40 shows the contour map for rotations about the $C_4'$-$C_5'$ and $C_5'$-$O_5'$ bonds of 2' endo ribose. Since the 2' OH is far from $C_5'$ we expect the analogous energy surface for the deoxyribose compound to be the same for all intents and purposes. The contour in the middle of the figure which has the shape of a dog-bone surrounds the region of lowest energy. The regions of high energy are for orientations of $C_5'$-$O_5'$ near 0° and the orientations of $C_4'$-$C_5'$ near 0°, 240° and to a lesser extent, 120°. These are caused by repulsive steric interactions of the phosphate phosphorus with atoms of the ribose group. At 0° orientation of the $C_5'$-$O_5'$ bond a rotation of the $C_4'$-$C_5'$ bond brings the phosphorus atom closest to $O_1'$ at approximately 0°, $H_4'$ at approximately 120° and $H_3'$ at approximately 240°. These interactions can be seen easily in space filling models. The surface is generally flat with a few steric repulsions causing high energy regions. More extensive calculations of sterically allowed conformations have been made by Sasisekharan _et al_.[166]. Conformations are forbidden if interatomic distances become smaller than the sum of the hard core radii for the two atoms. In their analysis the allowed regions are approximately $C_5'$-$O_5'$ orientations from 150° to 210° and $C_4'$-$C_5'$ orientations from 50° to 80°, 140° to 200° and

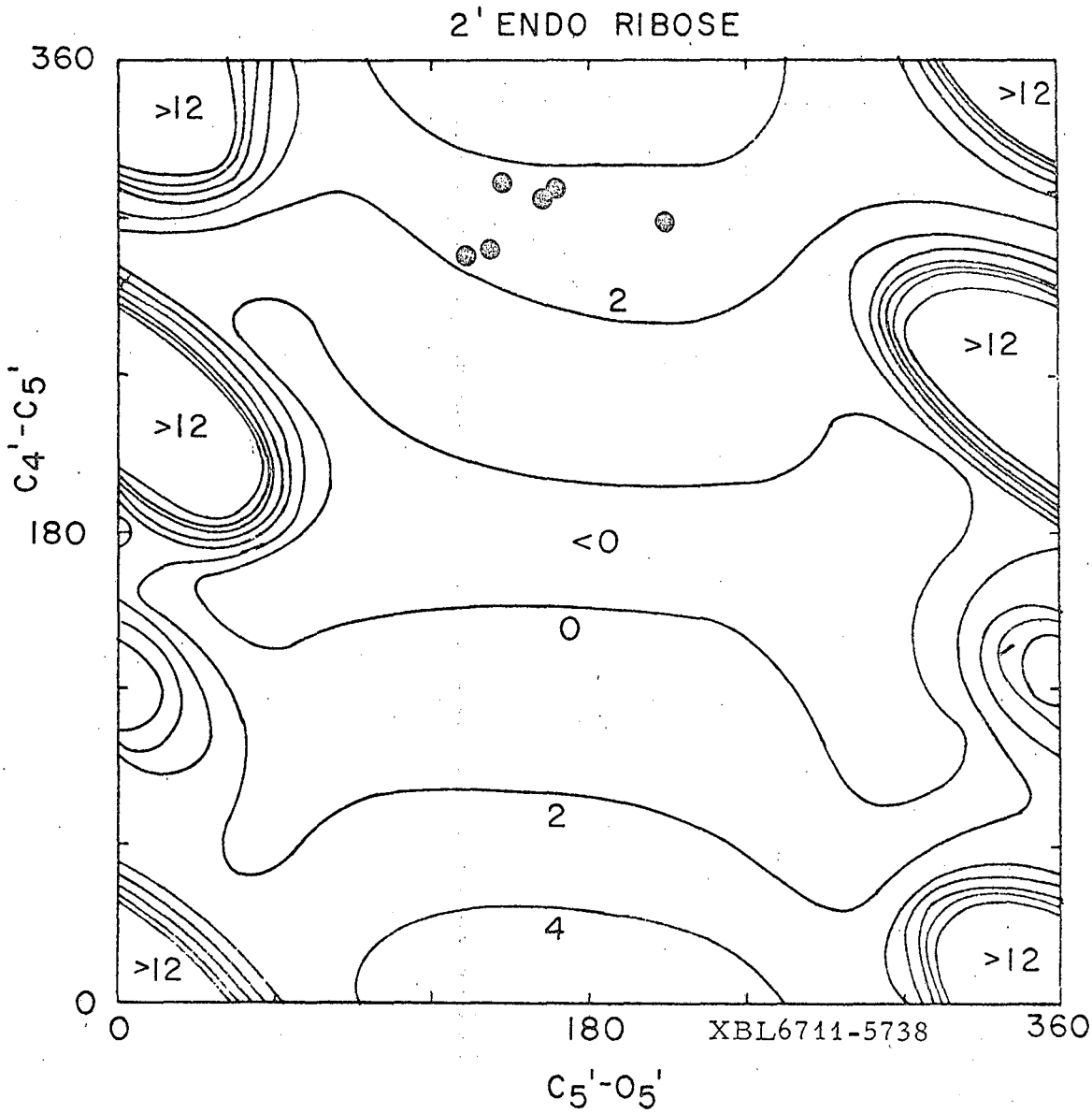



Figure 40 Calculated energy as a function of the $C_4'-C_5'$ and $C_5'-O_5'$ torsional angles for 2' endo ribose. Contours are at every 2 kcal per mole. The circles indicate polynucleotide structures.

280° to 310°. These are compatible with our calculations, although they eliminate far more area than we are able to on the basis of our calculations.

Torsional angles that are found in fiber structures of nucleic acids (by X-ray diffraction techniques) can also be compared with our results. These angles were determined by taking the coordinates of the proposed structures, then using a computer program, converting the cartesian coordinates to a cylindrical coordinate system, where the z axis is in the direction of the bond in question. The difference in the cylindrical angles, $\theta$, of the two atoms used to define the orientation of the bond, was the torsional angle for that bond. The orientation angles for some polynucleotide structures are listed in Table 19. They are indicated on the energy contour diagrams as circles.

It is somewhat surprising and a little discouraging to find that the torsional angles of these structures are not near the energy minimum which we calculated. It should be remembered that this region is only about 3 kcal per mole higher than the lowest energy found for this particular calculation. Since we have not included all possible interactions, those which are dependent on the orientations of bonds other than $C_4'$-$C_5'$ and $C_5'$-$O_5'$ (e.g. base-base interactions) could easily make a 3 kcal per mole difference. We will see later that this is the case in our calculations of the energies of the conformations of CpC.

The energy surface for rotations about the $C_3'$-$O_3'$ and $O_3'$-P bonds is quite different. (See Figures 41 and 42.)

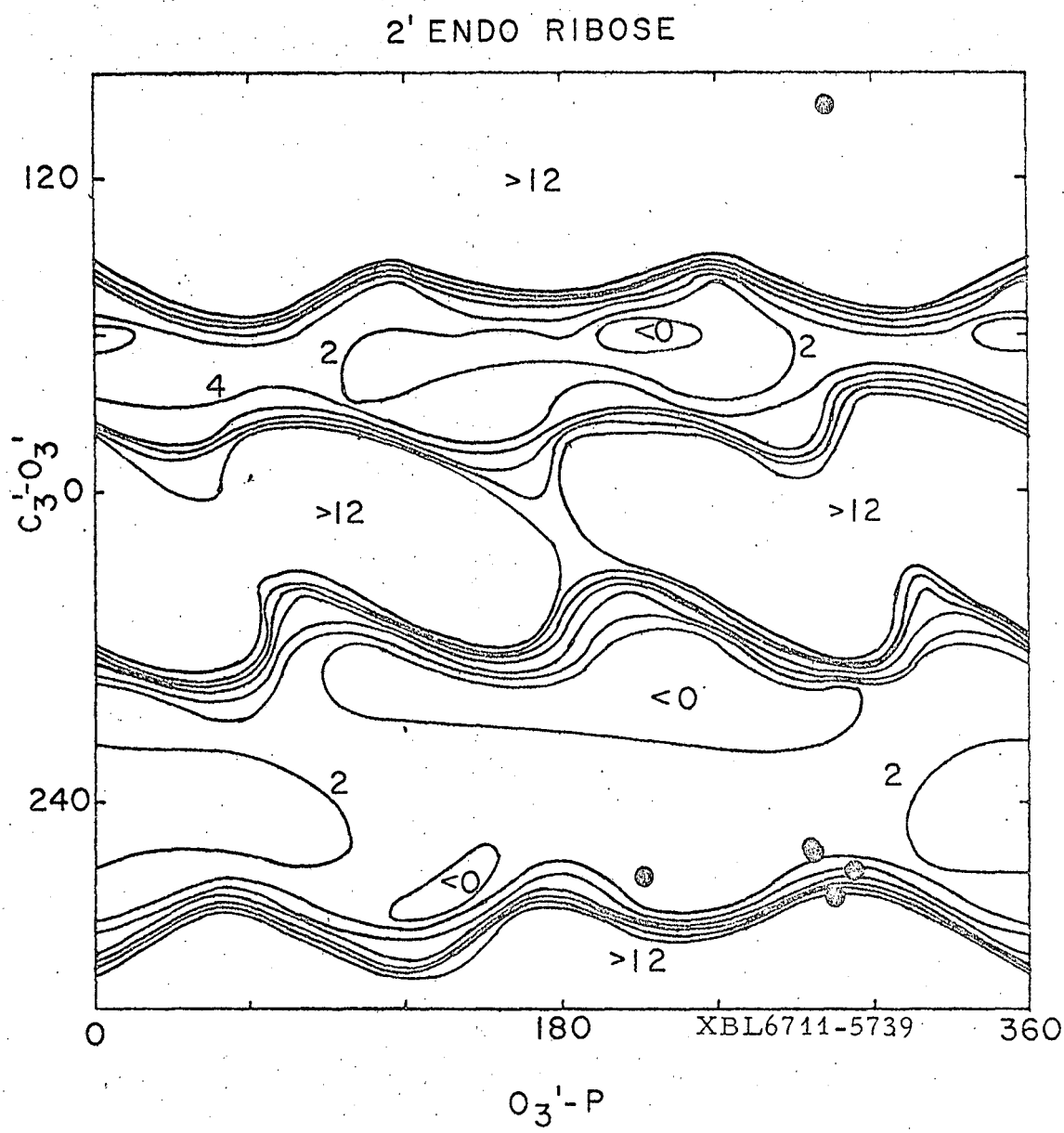



Figure 41 Calculated energy as a function of the $C_3'$-$O_3'$ and $O_3'$-P torsional angles for 2' endo ribose. Contours are at every 2 kcal per mole. The circles indicate polynucleotide structure.

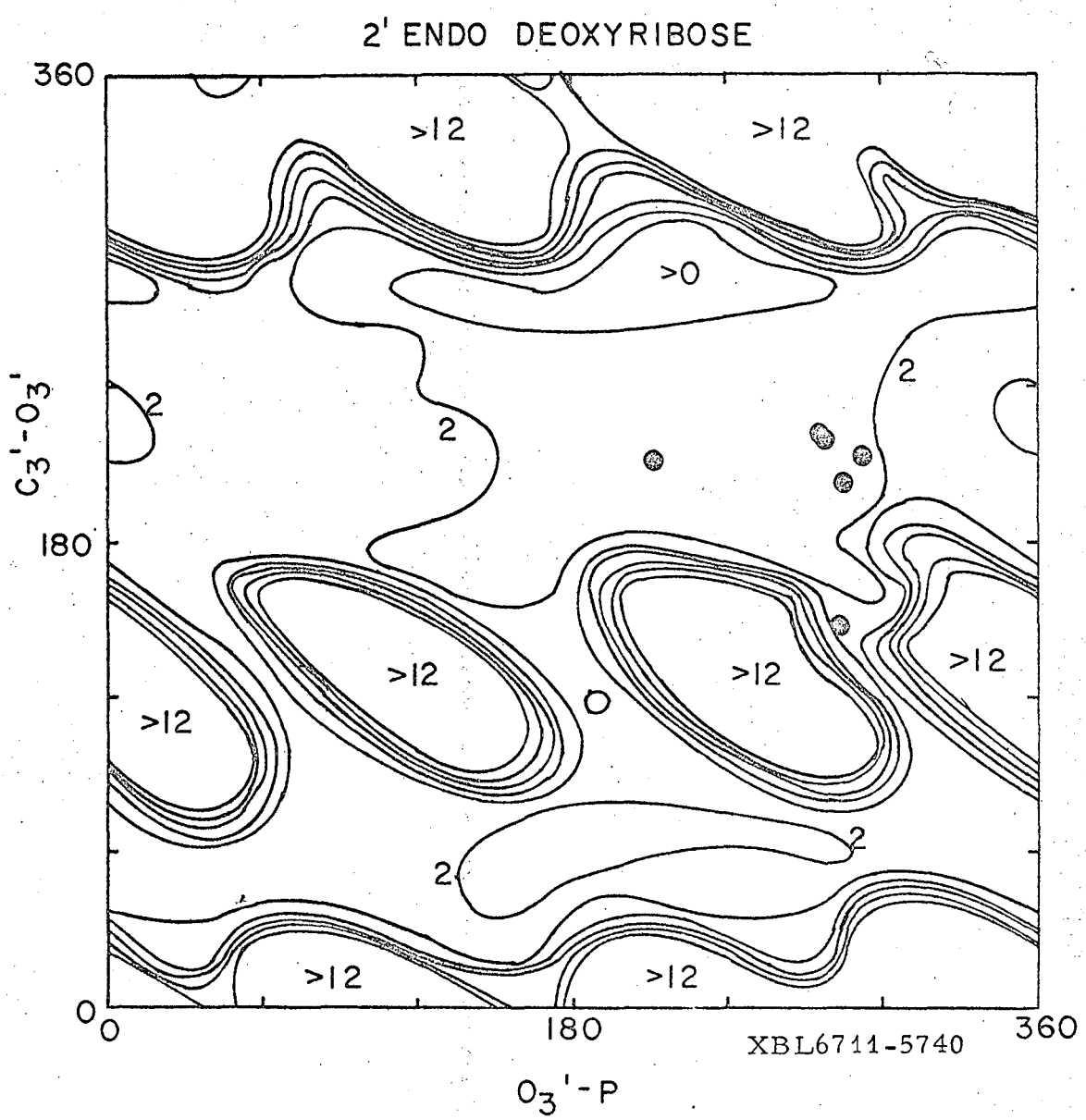



Figure 42 Calculated energy as a function of the $C_3'-O_3'$ and $O_3'-P$ torsional angles for 2' endo deoxyribose. Contours are at every 2 kcal per mole. The circles indicate polynucleotide structure.

TABLE 19

TORSIONAL ANGLES OF KNOWN STRUCTURES

| Structure | $\phi_{CN}$ | $C_3'-O_3'$ | $O_3'-P$ | $P-O_5'$ | $O_5'-C_5'$ | $C_5'-C_4'$ | Reference |
|---|---|---|---|---|---|---|---|
| A DNA | -14 | 221 | 279 | 282 | 168 | 312 | 63 |
| B DNA | -86 | 147 | 282 | 283 | 211 | 299 | 104 |
| C DNA | -73 | 211 | 212 | 315 | 143 | 289 | 122 |
| poly A (acid form) | 1 | 216 | 293 | 285 | 168 | 314 | 161 |
| RNA 10 residues/turn | - 5 | 203 | 285 | 341 | 134 | 287 | 62, 178 |
| RNA 11 residues/turn | -11 | 223 | 277 | 284 | 163 | 309 | 62, 178 |

Unlike the $C_4'$-$C_5'$ system, rotations are quite hindered. Instead of being a flat surface with a few forbidden areas, it is a surface which has much high energy area and low energy valleys with reasonably steep slopes. Because many of these close contacts involve the 2' OH group, there are considerable differences between the ribose and deoxyribose energy surfaces, the latter having considerably less forbidden area. In both cases there are two bands of high energy conformations across the contour diagram. One band, centered at a $C_3'$-$O_3'$ torsional angle of approximately 350°, results from interactions of the phosphate oxygens with the $H_4'$. The large forbidden region, (in the case of ribose) centered at a $C_3'$-$O_3'$ torsional angle of approximately 130°, results from the steric and charge-charge interactions of the phosphate oxygens with the 2' OH group. This region is greatly reduced in the deoxyribose sugar where a $H'_2'$ is substituted for the 2' OH. The regular oscillations of the edges of both forbidden regions (with a 120° period) result from interactions with the three phosphate oxygens. As rotation occurs about the $O_3'$-P bond, first one phosphate oxygen reaches its position of minimum distance from the ribose, then the second and finally the third.

The earlier work on the sterically forbidden regions (calculated on the basis of hard core repulsions only) predicts that the only orientations about the $C_3'$-$O_3'$ bond allowed should be approximately 260° to 270°.[166]. Our calculations are compatible with this result. The

allowed region is very close to the lowest energy area on the two dimensional energy surface. However, none of the structures determined experimentally (by fiber diffraction studies) fall in this region. This illustrates the point that a great deal of care must be taken in hard core repulsion calculations.

It should be noted that the orientation of the 2' OH used in our calculations is that found in Cp. It is possible for rotation to take place around the $C_2'$-$O_2'$ bond. That means that the steep increase in energy close to the $C_3'$-$O_3'$ orientation of 190° may move up or down as rotation takes place about the $C_2'$-$O_2'$ bond and the 2' OH is moved. Therefore we can not say that the 11 residue per turn structure proposed for double-strand ribosomal RNA is much more favorable than the 10 residue per turn structure. If there were a rotation about the $C_2'$-$O_2'$ bond and/or small changes in the geometry assumed for the furanose ring, the position of the 10 residue per turn structure may no longer be in an excessively high energy region.

The torsional angles for $C_3'$-$O_3'$ which are found in polynucleotide fiber structures are all in regions of reasonably low energy on the deoxyribose surface (See Figure 42.) As was the case with the $C_4'$-$C_5'$ surface they are not at the minimum. The energy difference between the polynucleotide coordinates and the energy minimum is only about 2 kcal per mole. The B form of DNA is in a strongly forbidden region of the energy surface in that case where ribose is the sugar (Figure 41). Using

Cortauld space filling molecular models an RNA oligomer can be put into a structure very close to that of the B form of DNA. Unfortunately, space filling models do not indicate coulombic repulsions. In this structure $O_2'$, a phosphate oxygen and $O_1'$ are very close to one another. A structure of double-strand RNA similar to the B form of DNA would be energetically unfavorable because each of these atoms has a charge equivalent to approximately 0.5 electron, causing a very strong coulombic repulsion.

It is surprising that amid all the discussion of differences between ribo and deoxyribopolynucleotides no one has mentioned this interaction as a possible origin of differences between these compounds. Exactly what effect this will have on the structures of polynucleotides is hard to say. It may cause the low temperature limit (ordered form) of single-strand ribo- and deoxyribopolynucleotide structures to be quite different. This will affect both the low temperature limit of the optical properties as well as the thermal disordering process. Such differences in the disordering ("melting") process have been observed for ribo- and deoxyribooligonucleotides and polynucleotides[197,206].

The next calculation involved is the elimination of additional orientations of the internucleoside bonds of such high energy that they do not have to be considered for further calculations. If half of the angular orientations of each bond could be eliminated, then calculations of the conformations of a dinucleoside phosphate (where there

are 8 bonds about which rotation may take place), could be reduced by a factor of 256. We therefore considered larger parts of the sugar-phosphate backbone. Calculations on a 30° grid (12 orientations per bond) were done in two cases. In the first case we calculated energies of conformations which were generated by rotations about the $O_3'$-P, P-$O_5'$, $O_5'$-$C_5'$, and $C_5'$-$C_4'$ bonds in which interactions were considered among the following atoms: $C_3'$, $O_3'$, P, phosphate O, phosphate O', $O_5'$, $C_5'$, $H_5'$, $H'_5'$, $C_4'$, $H_4'$, $O_1'$, $H_3'$, $O_3'$, and 3' hydroxyl H. In the second case we calculated energies of conformations generated by rotations about $C_3'$-$O_3'$, $O_3'$-P, P-$O_5'$ and $O_5'$-$C_5'$ in which interactions were considered among the following atoms: $C_3'$, $O_2'$, 2' hydroxyl H, $H_2'$, $H_3'$, $H_4'$, $H_5'$, $H'_5'$, $O_5'$, 5' hydroxyl H, $O_3'$, P, phosphate O, phosphate O', $O_5'$, $C_5'$, $H_5'$, $H'_5'$ and $C_4'$. The results of these calculations are presented in Figures 43-46. These results were analyzed to minimize the chance of discarding a conformation which may (including all possible interactions) have a low energy. The results are presented as two dimensional surfaces (two consecutive torsional angles, $C_3'$-$O_3'$ and $O_3'$-P for example). The surface was defined by assigning to each point the lowest energy for those two angular orientations and all possible orientations for the two other bonds considered in the calculation. Those points with energies greater than 20 kcal per mole more than the lowest energy were discarded. Unfortunately we were not able to eliminate very many

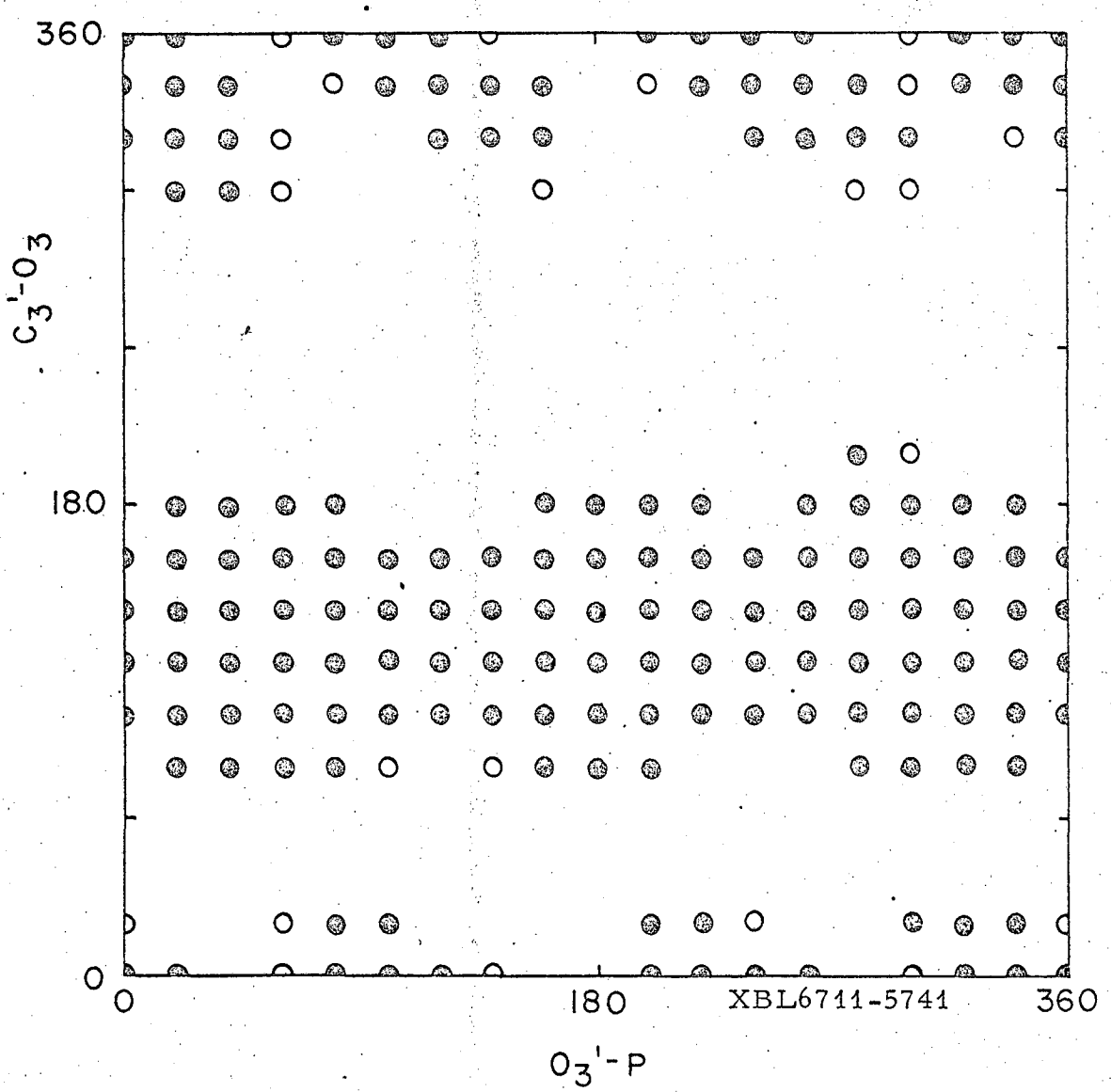



Figure 43 Conformations eliminated on the $C_3'-O_3'$, $O_3'-P$ surface. The open circles indicate energies between 20 and 50 kcal per mole above the energy minimum. Filled circles indicate energies more than 50 kcal per mole above the energy minimum.

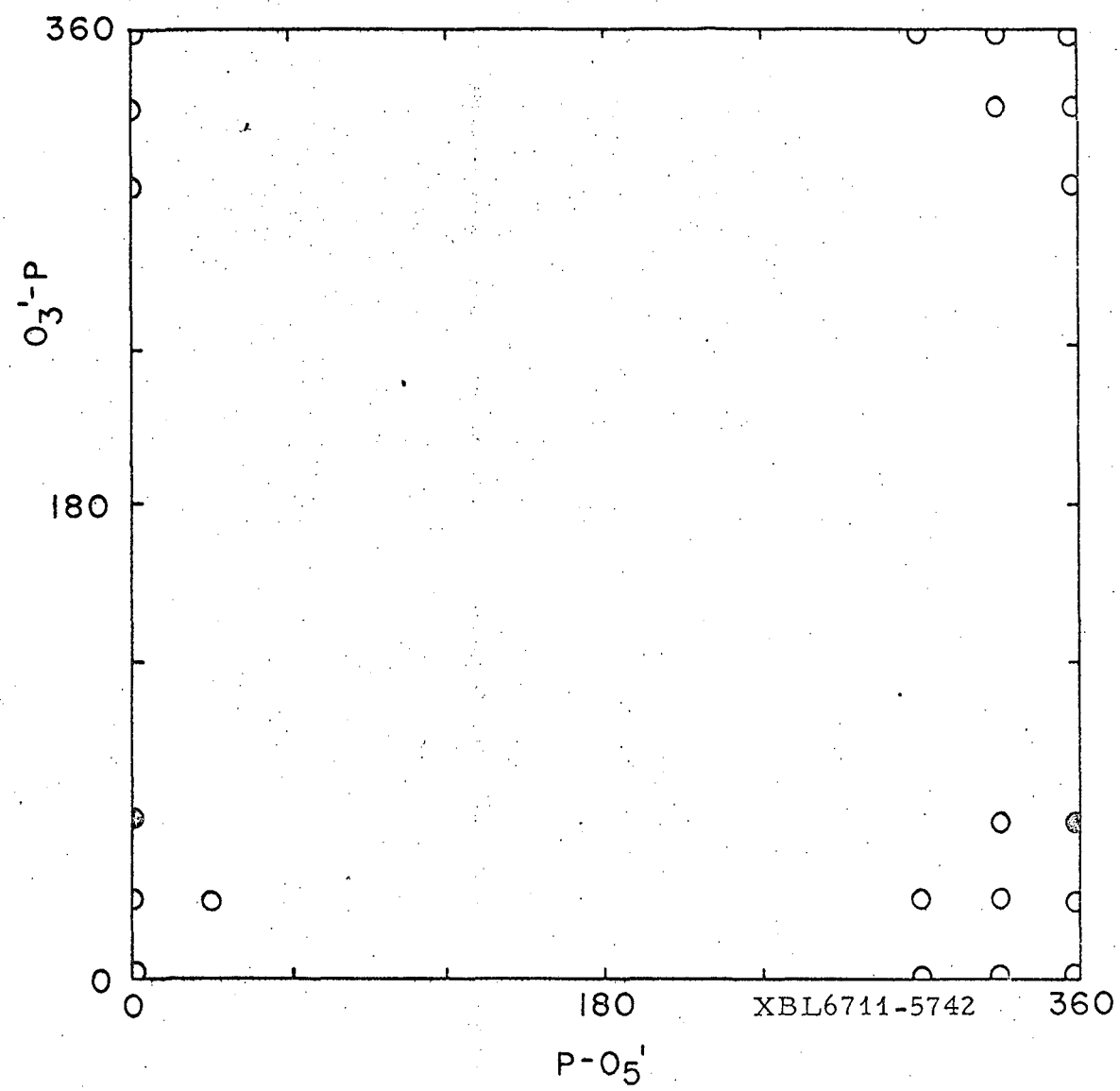



Figure 44 Conformations eliminated on the $O_3'$-P, P-$O_5'$ surface. The open circles indicate energies between 20 and 50 kcal per mole above the energy minimum. Filled circles indicate energies more than 50 kcal per mole above the energy minimum.

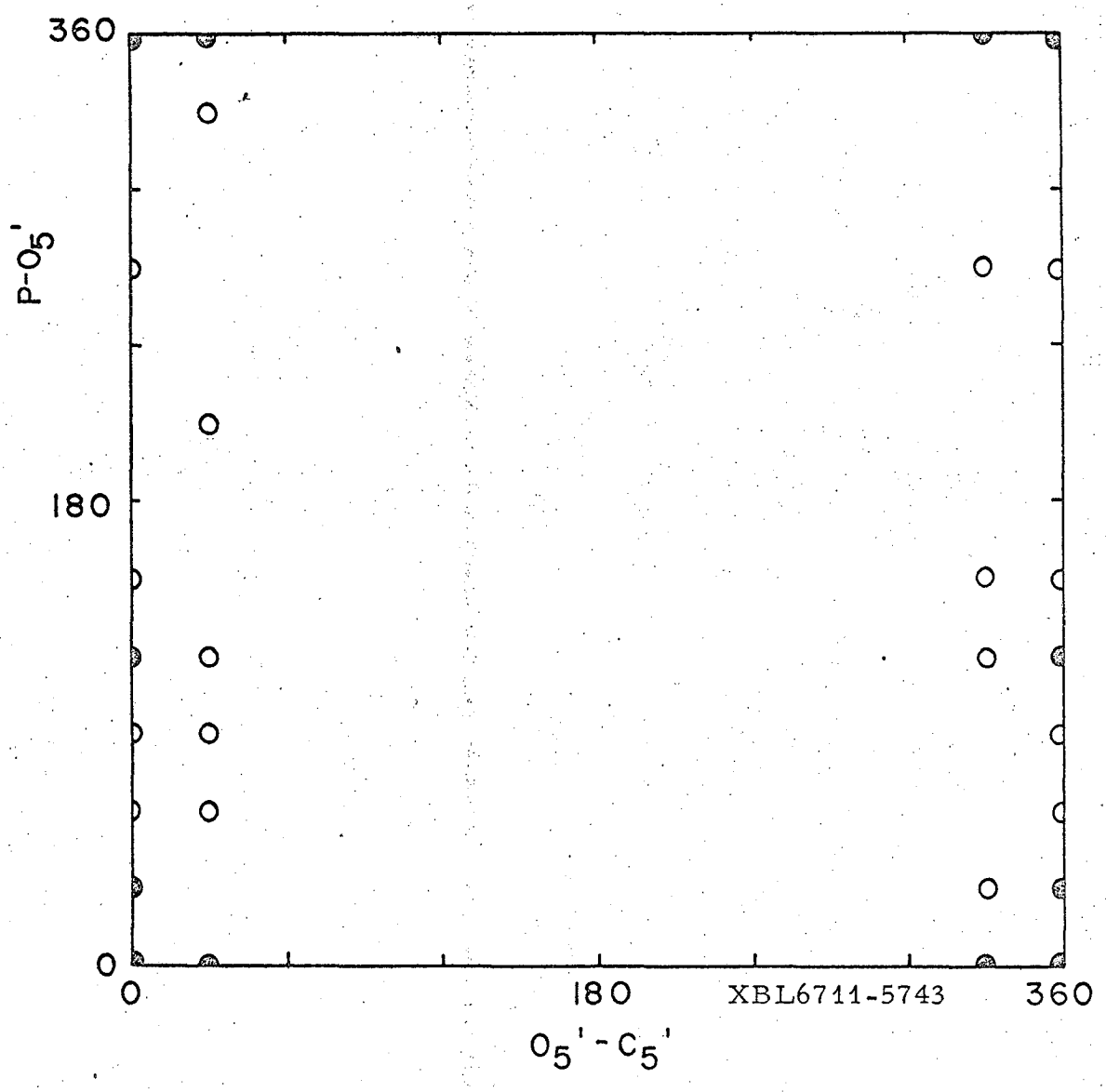



Figure 45 Conformations eliminated on the P-$O_5$'-$O_5$'-$C_5$' surface. The open circles indicate energies between 20 and 50 kcal per mole above the energy minimum. Filled circles indicate energies more than 50 kcal per mole above the energy minimum.

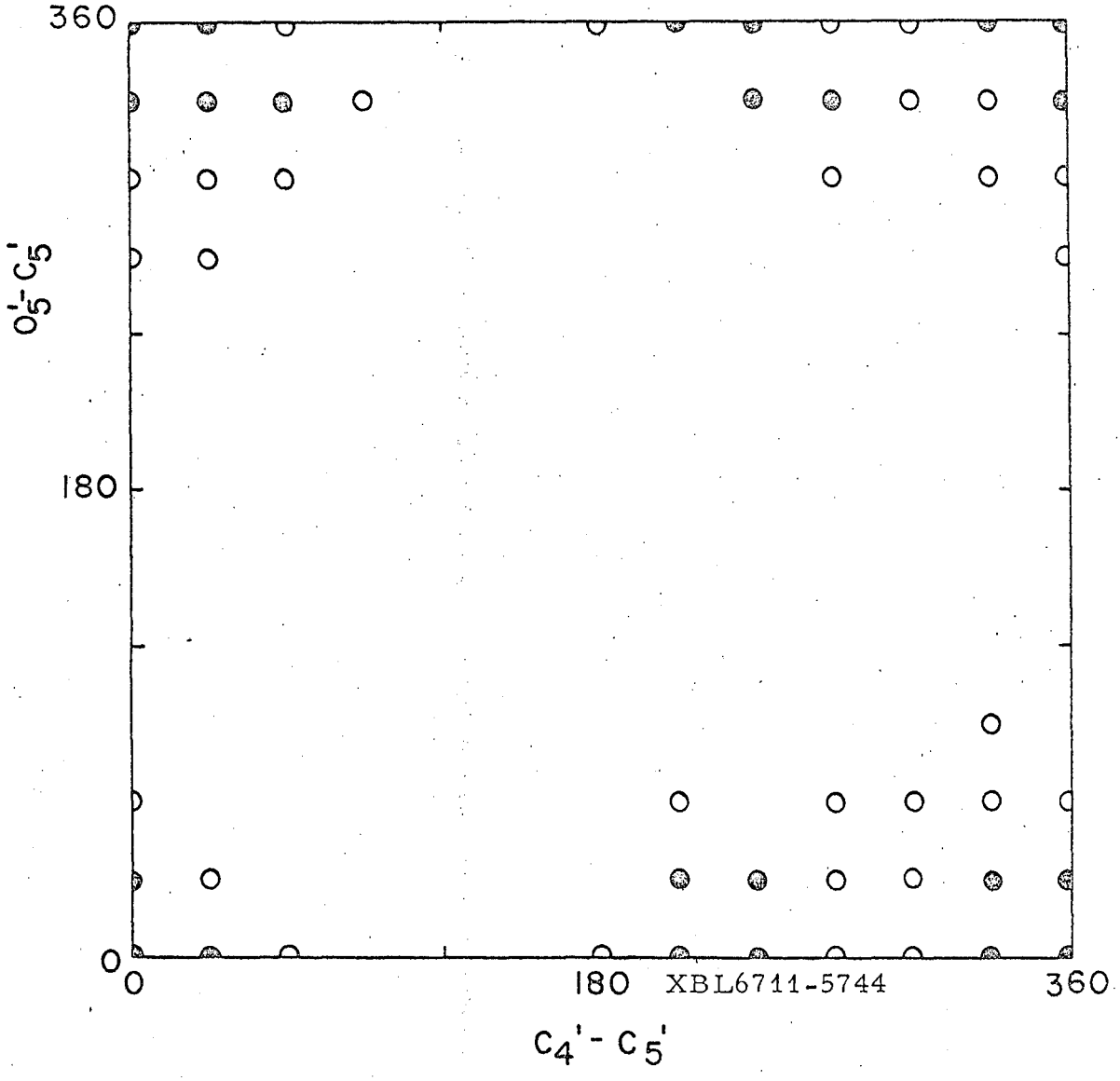



Figure 46 Conformations eliminated on the $O_5'-C_5'$, $C_5'-C_4'$ surface. The open circles indicate energies between 20 and 50 kcal per mole above the energy minimum. Filled circles indicate energies more than 50 kcal per mole above the energy minimum.

conformations. The $C_3'$-$O_3'$ surface had the highest fraction of conformations which we could discard, 49%. Most had a far smaller fraction.

CpC

The next more complicated structure to be calculated was a dinucleoside phosphate. CpC was chosen as the dinucleoside phosphate to be studied for a number of reasons. First of all, the two bases are the same. This symmetry simplifies the calculation somewhat and enables us to make some comparisons between the 3' and 5' nucleosides. If the bases were not the same, we would not be sure if any differences were caused by the difference in base or the difference in location of the phosphate. Since we know more about the orientation about the glycosidic bond for pyrimidine nucleosides (where only the small anti region need be considered) than we do about purine nucleosides (where all orientations would probably have to be treated) the choice of a dinucleoside phosphate with two pyrimidine bases is advantageous. Also, the best X-ray diffraction data available for nucleosides and nucleotides are those for Cp. Two such studies have recently been reported for the orthorhombic and monoclinic forms of Cp[23,182]. Although the crystal densities and intermolecular hydrogen bonding schemes are quite different in the two cases, the molecular conformations are almost identical. The only appreciable difference between the two structures is a rotation about the $O_3'$-P bond, a change that does not affect our calculations.

Because the conformation of the molecule is so similar in two different environments we hope that it is the same in a third environment, part of a dinucleoside phosphate in aqueous solution. That is a very big extrapolation to make. (The furanose conformation in cytidine is 3' endo[58].) However, we feel that our chances of choosing the appropriate crystal structure for our calculations are best for CpC.

The first attempt we made to find the conformations of CpC corresponding to the lowest energies, was to orient the dinucleoside phosphate in a conformation which seemed reasonable (a conformation found in polynucleotides by fiber diffraction techniques, or a structure which seemed most promising on the basis of model building with space filling molecular models) and then to allow it to relax to lower energies. Starting from a large number of conformations, if only a few minima were found, we could be reasonably confident that we were not missing any conformations of low energies.

Two types of minimization routines were tried. One of them (FINDMIN) took the molecule at the starting conformation and systematically calculated energies for torsional angles (one bond at a time) a specified increment larger and then smaller than the torsional angle for that bond in the starting conformation. If it found a lower energy, it changed the conformation of CpC to the new lower energy conformation and continued trying other bonds. When no lower energy conformation could be found (i.e. the routine thought it had reached an energy minimum) the

computation stopped. The other routine used was based on Davidon's variable metric technique[41]. (The program, VARMIT, written by Eric Beals of the UCLRL Computing Center, is in that computing center's library.) Neither technique was successful. Invariably the routines were trapped in local minima very rapidly. (The lowest minimum found here was 17 kcal per mole higher than the lowest one we were eventually able to find.) There was almost a 1:1 correspondence of starting conformations and local minima found. That meant that we had no idea whether other choices of starting conformations would lead to lower energy conformations. Similar techniques have met with the same type of failure in calculations on oligopeptides[65]. In that case, the approach giving the most success was Davidon's variable metric technique, one of the two we tried. Since we had no success using the technique of direct minimization of energy, starting with conformations apparently having low energies we abandoned this meretricious short-cut and proceeded by other means.

The approach taken towards finding the lowest energy conformations of CpC was a brute force calculation of the energies of essentially all possible conformations. (BRUTE) We tried to eliminate as many unnecessary calculations as possible in two ways. The first way was not to calculate the energy of any conformation which was shown to have an unreasonably high energy by the extended calculations on the orientation of the internucleoside bonds, etc. Because we were convinced that C could only exist in the anti conformation we chose only values of $\phi_{CN}$ in that region.

Earlier minimization attempts gave $C_2'-O_2'$ orientation angles of 230° to 350° so that only orientation angles in that range were considered. Operationally this was done either by not introducing torsional angles which guaranteed high energies (e.g. for any torsional angles for the other 7 bonds a $C_3'-O_3'$ orientation of 120° would lead to a high energy) or by sorting through the sets of angles in the program (Subroutine SORT) and discarding those sets of torsional angles known to have unfavorable energies. (As listed in Appendix B, SORT eliminates orientations for CpC.) The second criterion for discarding a conformation was a short interatomic distance. If an interatomic distance less than 65% of the sum of the Van der Waals radii was found, then the conformation was discarded immediately. Interactions were calculated in an order such that small interatomic distances tended to occur early.) An interatomic distance that short gives rise to a strong $r^{-12}$ repulsive energy term which is at least 20 kcal per mole in the case of H-H interactions and considerably greater in other cases. The orientation angles tried are listed in Table 20. They comprise approximately a 40° grid.

TABLE 20

| Bond | Torsional angles tried |
|---|---|
| Glycosidic | -20°, -60° |
| $C_2'-O_2'$ | 260°, 300°, 340° |
| $C_3'-O_3'$ | 20°, 60°, 210°, 250°, 290°, 340° |
| $O_3'-P$ | 0°, 40°, 80°, 120°, 160°, 200°, 240°, 280°, 320° |
| $P-O_5'$ | 0°, 40°, 80°, 120°, 160°, 200°, 240°, 280°, 320° |
| $O_5'-C_5'$ | 0°, 40°, 80°, 120°, 160°, 200°, 240°, 280°, 320° |
| $C_5'-C_4'$ | 0°, 40°, 80°, 120°, 160°, 200°, 240°, 280°, 320° |

In all 121,161 different conformations were calculated. This required approximately 9 hours of central processing time on the CDC 6600's at the Lawrence Radiation Laboratory Computing Center. The lowest energy conformations which were found are listed in Table 21. The energy scale is arbitrary to an additive constant.

Starting with these orientations, minimization of the energy was performed. In order to save time, minimization (using FINDMIN) was done first with 10° steps, then with 5° steps. The result of these minimizations are presented in Table 22.

Although a large number of local minima were found, many of them have similar torsional angles and are grouped together. (Conformations number 1 and 3 of Table 22 have torsional angles which differ by no more than 5° for any bond.) Within 9 kcal per mole of the lowest energy found there are 18 local minima which fall into 6 different groups. Since the molecular geometries of conformations in the same group are so similar we need only discuss one member

TABLE 21

TORSIONAL ANGLES OF LOW ENERGY CONFORMATIONS ON THE 40° GRID

| CONFORMATION | 3'GLYCOSIDIC | | | | | | | 5'GLYCOSIDIC | ENERGY |
|---|---|---|---|---|---|---|---|---|---|
| NUMBER | $\phi_{CN}$ | $C_2'-O_2'$ | $C_3'-O_3'$ | $O_3'-P$ | $P-O_5'$ | $O_5'-C_5'$ | $C_5'-C_4'$ | $\phi_{CN}$ | (KCAL PER MOLE) |
| 1 | -20 | 340 | 290 | 120 | 120 | 240 | 320 | -20 | -122.4 |
| 2 | -20 | 300 | 290 | 120 | 120 | 240 | 320 | -20 | -121.2 |
| 3 | -20 | 300 | 290 | 280 | 160 | 240 | 320 | -20 | -120.8 |
| 4 | -20 | 340 | 250 | 240 | 240 | 120 | 280 | -20 | -120.6 |
| 5 | -20 | 300 | 290 | 280 | 160 | 240 | 320 | -20 | -120.3 |
| 6 | -20 | 260 | 290 | 280 | 160 | 240 | 320 | -20 | -119.2 |
| 7 | -20 | 260 | 290 | 120 | 120 | 240 | 320 | -20 | -118.9 |
| 8 | -20 | 300 | 210 | 0 | 240 | 160 | 240 | -20 | -118.3 |
| 9 | -20 | 260 | 210 | 240 | 240 | 120 | 280 | -20 | -117.4 |
| 10 | -20 | 260 | 210 | 0 | 240 | 160 | 240 | -20 | -117.4 |
| 11 | -20 | 300 | 210 | 320 | 160 | 240 | 280 | -20 | -117.1 |
| 12 | -60 | 260 | 290 | 280 | 160 | 240 | 320 | -20 | -117.0 |
| 13 | -60 | 300 | 290 | 280 | 160 | 240 | 320 | -20 | -116.7 |
| 14 | -20 | 300 | 210 | 40 | 200 | 160 | 280 | -20 | -116.7 |

TABLE 21, cont.

TORSIONAL ANGLES OF LOW ENERGY CONFORMATIONS ON THE 40° GRID

| CONFORMATION | 3'GLYCOSIDIC | | | | | | | 5'GLYCOSIDIC | ENERGY |
|---|---|---|---|---|---|---|---|---|---|
| NUMBER | $\phi_{CN}$ | $C_2'-O_2'$ | $C_3'-O_3'$ | $O_3'-P$ | $P-O_5'$ | $O_5'-C_5'$ | $C_5'-C_4'$ | $\phi_{CN}$ | (KCAL PER MOLE) |
| 15 | -60 | 300 | 210 | 40 | 200 | 160 | 280 | -20 | -116.4 |
| 16 | -20 | 340 | 290 | 280 | 240 | 120 | 280 | -20 | -116.4 |
| 17 | -20 | 300 | 210 | 240 | 240 | 120 | 280 | -20 | -115.9 |
| 18 | -20 | 340 | 210 | 0 | 240 | 160 | 240 | -20 | -115.6 |
| 19 | -20 | 340 | 290 | 120 | 160 | 240 | 320 | -20 | -115.5 |
| 20 | -20 | 260 | 250 | 240 | 240 | 120 | 280 | -20 | -115.3 |
| 21 | -20 | 300 | 290 | 120 | 160 | 240 | 320 | -20 | -115.1 |
| 22 | -60 | 300 | 250 | 200 | 280 | 120 | 240 | -60 | -115.1 |
| 23 | -20 | 260 | 250 | 160 | 40 | 240 | 0 | -20 | -115.1 |
| 24 | -60 | 260 | 250 | 160 | 40 | 240 | 0 | -20 | -115.0 |
| 25 | -20 | 340 | 290 | 240 | 160 | 240 | 320 | -20 | -115.0 |

TABLE 22

| CONFORMATION NUMBER | 3'NUCLEOSIDE $\phi_{CN}$ | $C_2'-O_2'$ | $C_3'-O_3'$ | $O_3'-P$ | $P-O_5'$ | $O_5'-C_5'$ | $C_5'-C_4'$ | 5'NUCLEOSIDE $\phi_{CN}$ | ENERGY (KCAL/MOLE) | GROUP |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -10 | 335 | 290 | 120 | 100 | 245 | 320 | -15 | -134.3 | I |
| 2* | -25 | 30 | 280 | 220 | 125 | 255 | 320 | -35 | -134.1 | II |
| 3 | -10 | 335 | 295 | 125 | 100 | 245 | 320 | -20 | -133.5 | I |
| 4* | -15 | 335 | 290 | 265 | 160 | 255 | 325 | - 5 | -133.2 | III |
| 5 | -15 | 340 | 295 | 115 | 105 | 240 | 320 | -15 | -132.7 | I |
| 6* | -25 | 35 | 275 | 230 | 130 | 250 | 320 | -35 | -132.4 | II |
| 7 | -15 | 335 | 280 | 285 | 160 | 245 | 320 | -15 | -132.1 | III |
| 8 | -15 | 335 | 275 | 285 | 165 | 245 | 320 | -10 | -131.6 | III |
| 9 | -20 | 40 | 280 | 290 | 240 | 120 | 280 | -30 | -131.4 | IV |
| 10 | -10 | 335 | 285 | 125 | 100 | 250 | 320 | -20 | -131.1 | I |
| 11 | -10 | 310 | 200 | 320 | 160 | 235 | 290 | - 5 | -131.1 | V |
| 12 | - 5 | 295 | 200 | 5 | 235 | 165 | 240 | -10 | -130.8 | VI |
| 13 | -15 | 335 | 275 | 290 | 165 | 240 | 320 | -10 | -130.3 | III |
| 14 | -25 | 40 | 270 | 235 | 135 | 245 | 320 | -30 | -130.2 | II |
| 15 | -25 | 5 | 285 | 275 | 240 | 115 | 280 | -25 | -130.2 | IV |

* Found in the search for a low energy path from conformation 1 to conformation 7.

TABLE 22

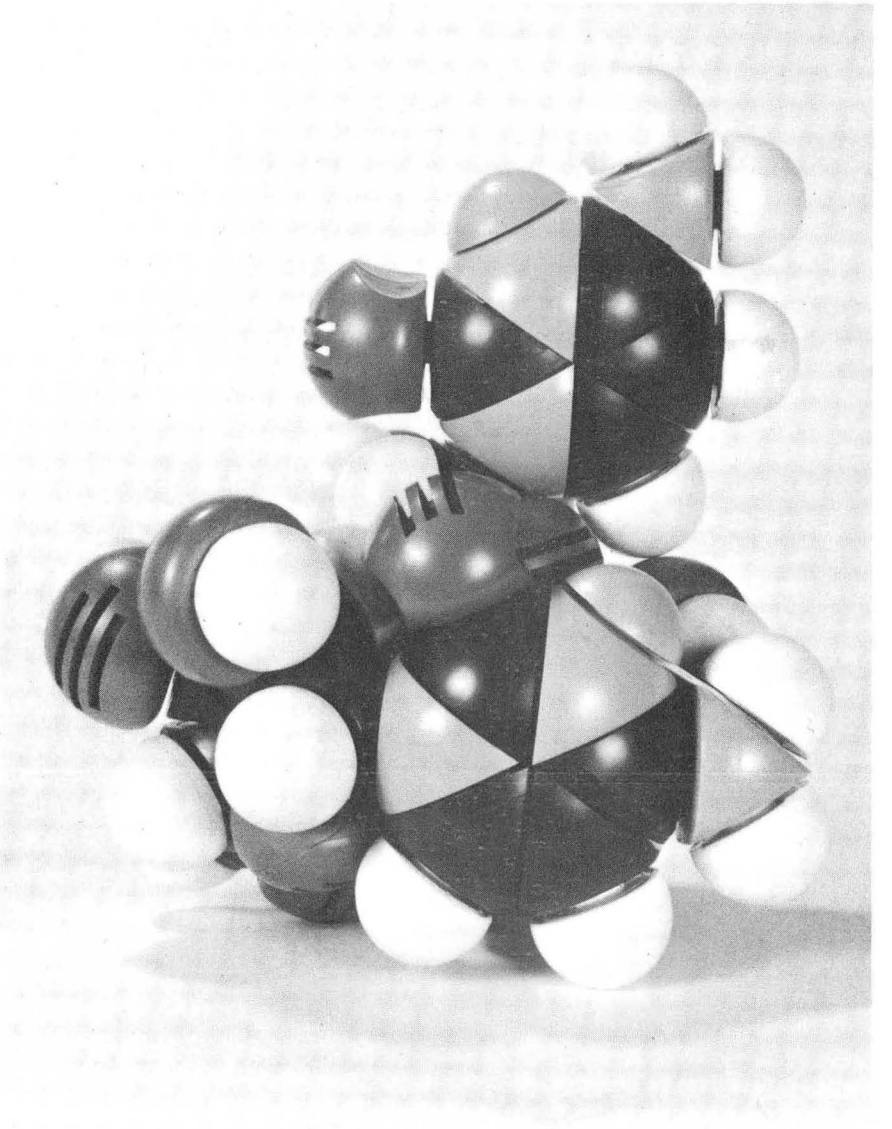
| CONFORMATION NUMBER | 3'NUCLEOSIDE $\phi_{CN}$ | $C_2'-O_2'$ | $C_3'-O_3'$ | $O_3'-P$ | $P-O_5'$ | $O_5'-C_5'$ | $C_5'-C_4'$ | 5'NUCLEOSIDE $\phi_{CN}$ | ENERGY (KCAL/MOLE) | GROUP |
|---|---|---|---|---|---|---|---|---|---|---|
| 16 | -25 | 30 | 270 | 240 | 130 | 245 | 320 | -30 | -129.5 | II |
| 17 | -20 | 335 | 285 | 135 | 115 | 250 | 320 | -25 | -128.2 | I |
| 18 | -50 | 285 | 205 | 5 | 235 | 160 | 240 | -10 | -127.7 | VI |
| 19 | -40 | 265 | 240 | 300 | 175 | 250 | 310 | -10 | -125.2 | VII |
| 20 | -35 | 250 | 200 | 240 | 250 | 105 | 280 | -10 | -125.0 | VIII |
| 21 | -55 | 285 | 210 | 250 | 205 | 160 | 280 | - 5 | -123.8 | IX |
| 22 | -50 | 280 | 260 | 200 | 280 | 120 | 240 | -50 | -122.9 | X |

of a group. Figures 47 and 48 are photographs of conformation number 1 constructed with Pauling-Corey-Koltun space-filling models. Figures 49-58 illustrate the lowest energy conformations of each of the ten groups we have found. The view is perpendicular to the plane of the 3' cytosine with the 5' nucleoside indicated by dashed lines.

It is very difficult for the experimentalist to measure torsional angles for CpC in solution. A more easily determined characteristic of a dinucleoside phosphate is the relative orientation of the two bases. (Major contributions to ORD, ultraviolet absorption and NMR originate in conformation-dependent base-base interactions.) Comparison of conformations with this criterion in mind show even more similarities. There are three major types of base-base orientations, with large differences between them.

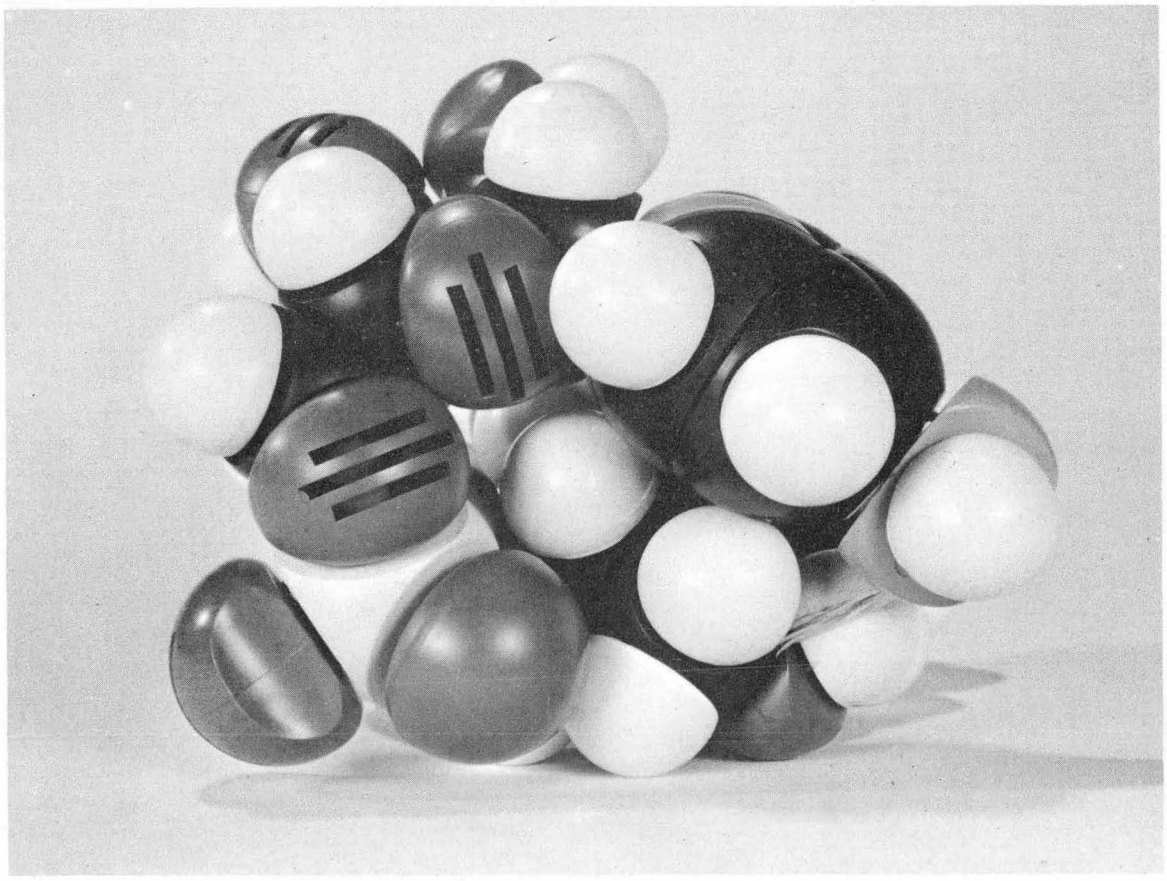
GROUP I

The conformations represented by group I are not what we expected as the most stable conformations of CpC. We expected the start of a right-handed helix (because all polynucleotide fiber structures determined up to this time have been right handed helices) with considerable overlapping of "stacked" bases. The two cytosine bases of this conformation are almost parallel to one another with left-handed helicity and almost no base-base overlap. (The orientation of the two bases is shown in Figure 47.) Instead, the cytosine of the 5' nucleoside is essentially the minimum contact distance from the 3' ribose, (See Figure 48) with the other



CBB 670-5957

Figure 47. The lowest energy conformation of CpC built with Corey-Pauling-Koltun space-filling models showing the orientation of the two bases.



CBB 670-5947

Figure 48. The lowest energy conformation of CpC built with Corey-Pauling-Koltun space-filling models showing the proximity of the 5' cytosine to the 3' ribose.

cytosine is almost completely exposed to the vacuum. Only very minor modifications of this structure are needed to extend it to form a polynucleotide. This helix would be very tight, with fewer than six (possibly as few as four) residues per turn. There is only one obvious specific interaction in this conformation. The 2' OH of the 5' ribose may form something resembling a hydrogen bond with the carbonyl oxygen of the other cytosine. However, this would be an unusual hydrogen bond. We will consider it in more detail when we discuss the role of hydrogen bonds in a single-strand polynucleotide structure.

GROUPS II AND IV

In conformations of groups II and IV there is almost no base-base overlap, but there is not enough room to fit a layer of solvent between the two bases. Because of this we still consider these conformations to be "stacked". The two cytosines are approximately perpendicular to each other, with the carbonyl of the 5' cytosine down and away from the 3' cytosine and the amino group of the 5' cytosine close to the carbonyl of the other base, approximately the minimum contact distance away. It is difficult to imagine a regular polynucleotide structure having nearest neighbor nucleotide interactions similar to the ones in these structures.

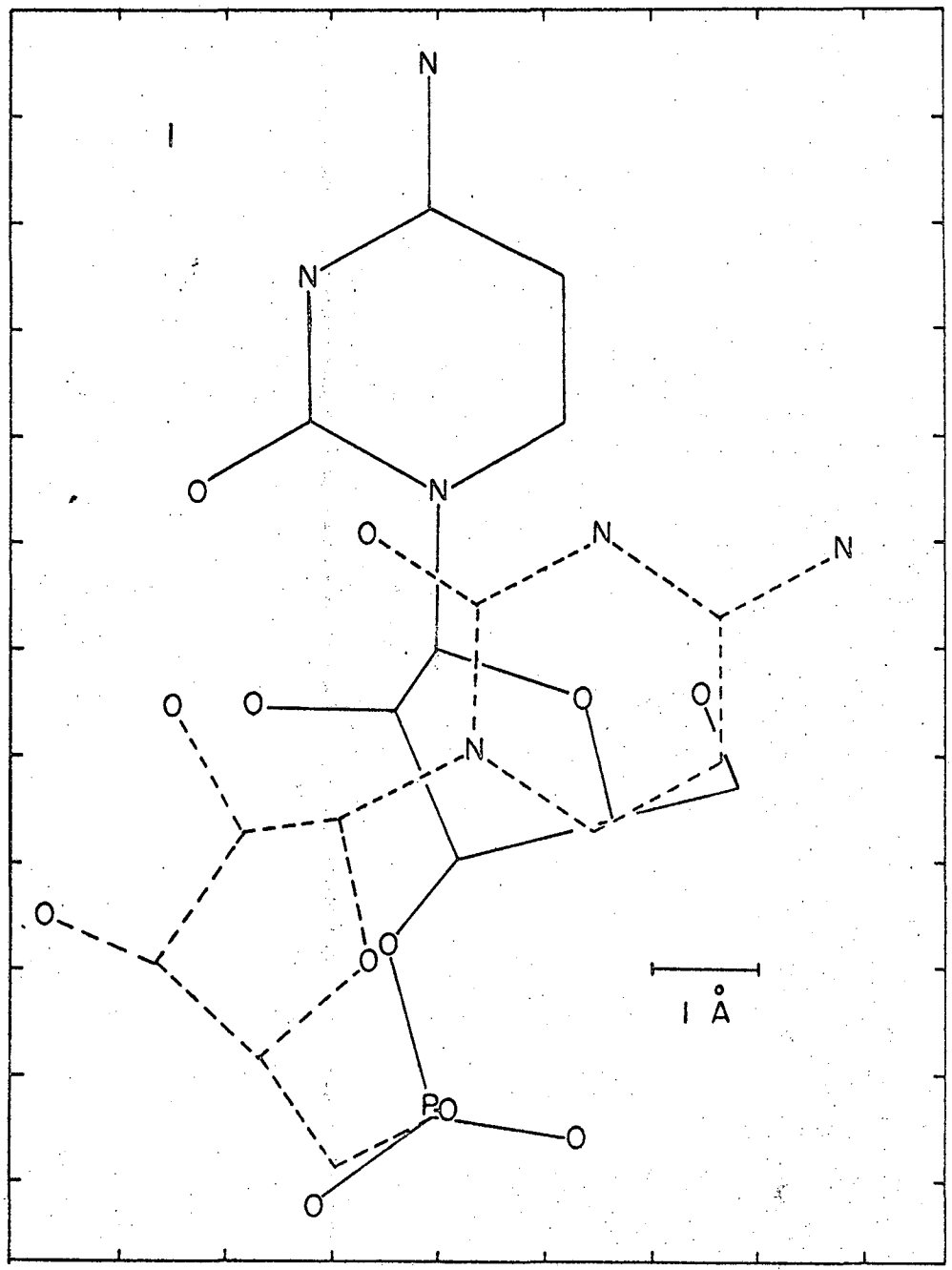
GROUPS III, V, VI, VII AND IX

The third type of conformation is found in groups III, V, VI, VII and IX. In these conformations the bases are

approximately parallel to one another with the 5' base about 4Å underneath the 3' base. The entire 5' nucleoside is rotated 180° about the glycosidic bond. The analogous parts of the bases (e.g. the carbonyls) are far from one another, rather than being close which would be the case if the bases had been oriented in the same manner. As was the case with the two other types of orientations already mentioned, there is very little base overlap with the 5' cytosine apparently interacting strongly with the 3' ribose.

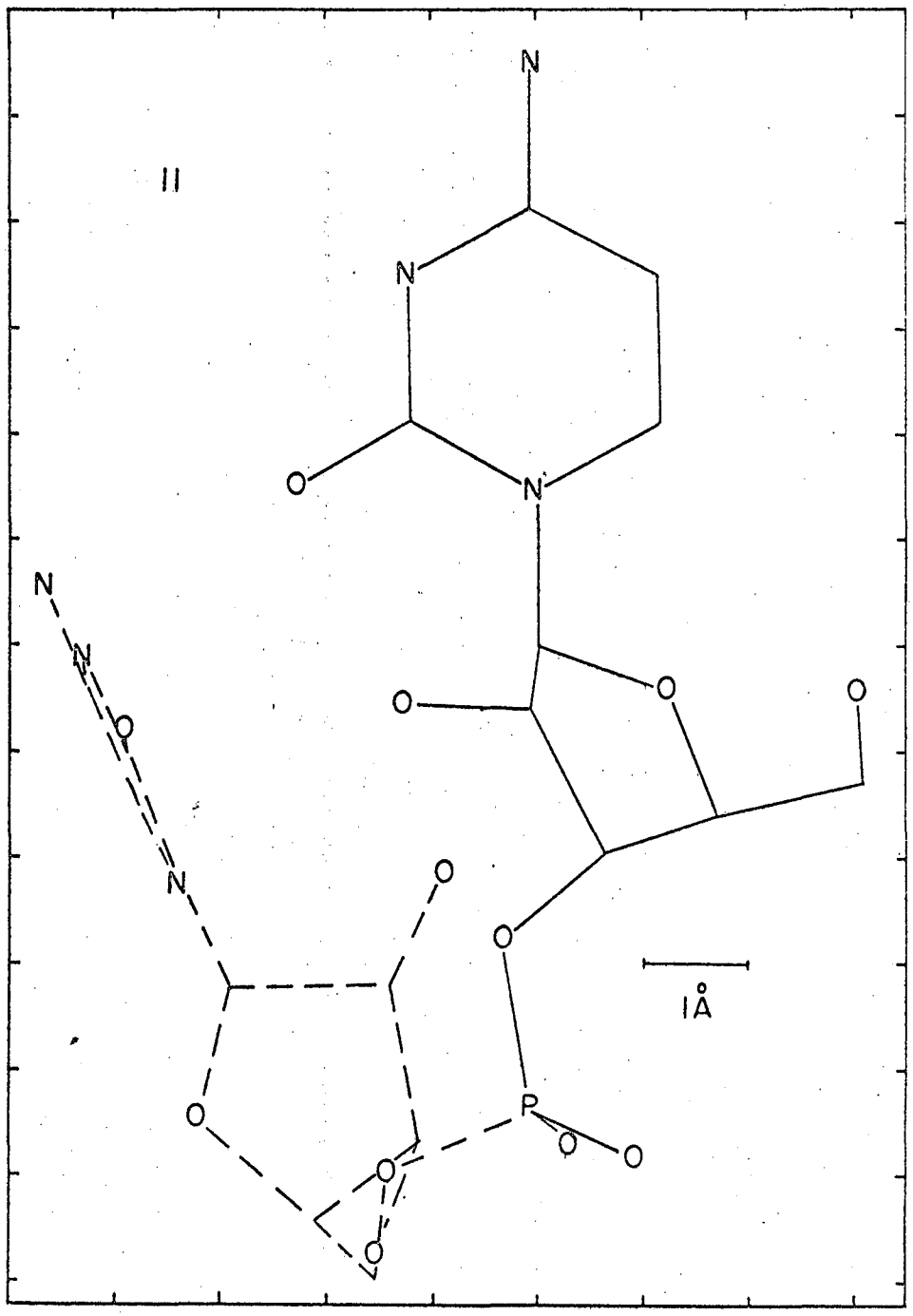
GROUP VIII

If we accept as our definition of an "unstacked" dinucleoside phosphate, one in which a layer of solvent does not completely isolate one base from the other then the 18 lowest energy conformations are "stacked". The conformation in group VIII is the only "unstacked" conformation we have found. Although one can build reasonable models of conformations where the bases are further apart, this conformation is definitely "unstacked" with the bases approximately 7Å away from one another, far enough to permit a layer of solvent to exist between them. The energy of this conformation is about 9 kcal per mole higher than the lowest energy conformation (having bases close to one another). This compares well with values of $\Delta H^{o}$ ("unstacking") of 4 to 10 kcal per mole determined from temperature dependent properties of dinucleoside phosphates. (These results are discussed at length in




XBL6711-5745

Figure 49 The lowest energy Group I conformation. The 5' nucleoside is indicated by dashed lines.




XBL6711-5746

Figure 50 The lowest energy Group II conformation. The 5' nucleoside is indicated by dashed lines.

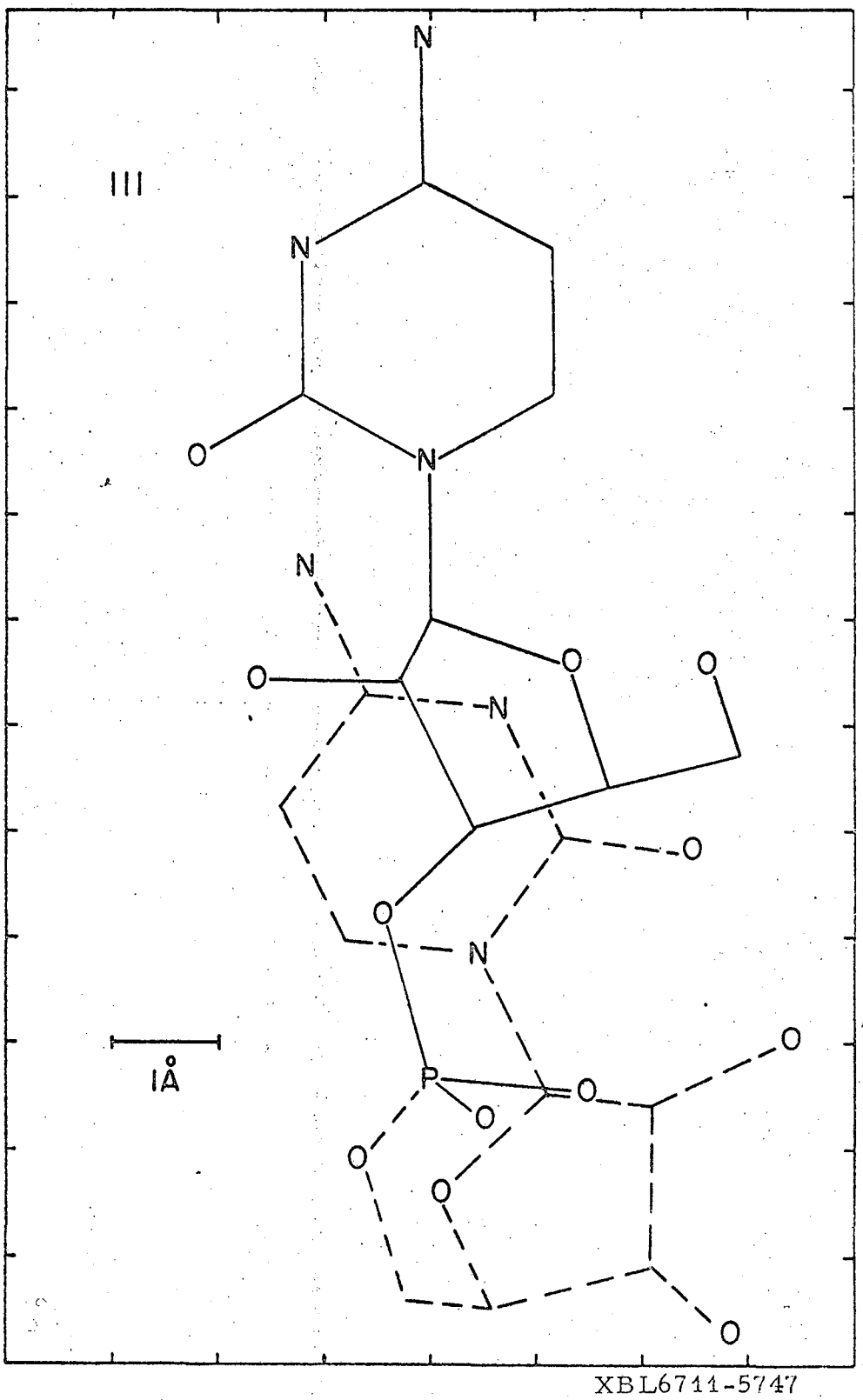
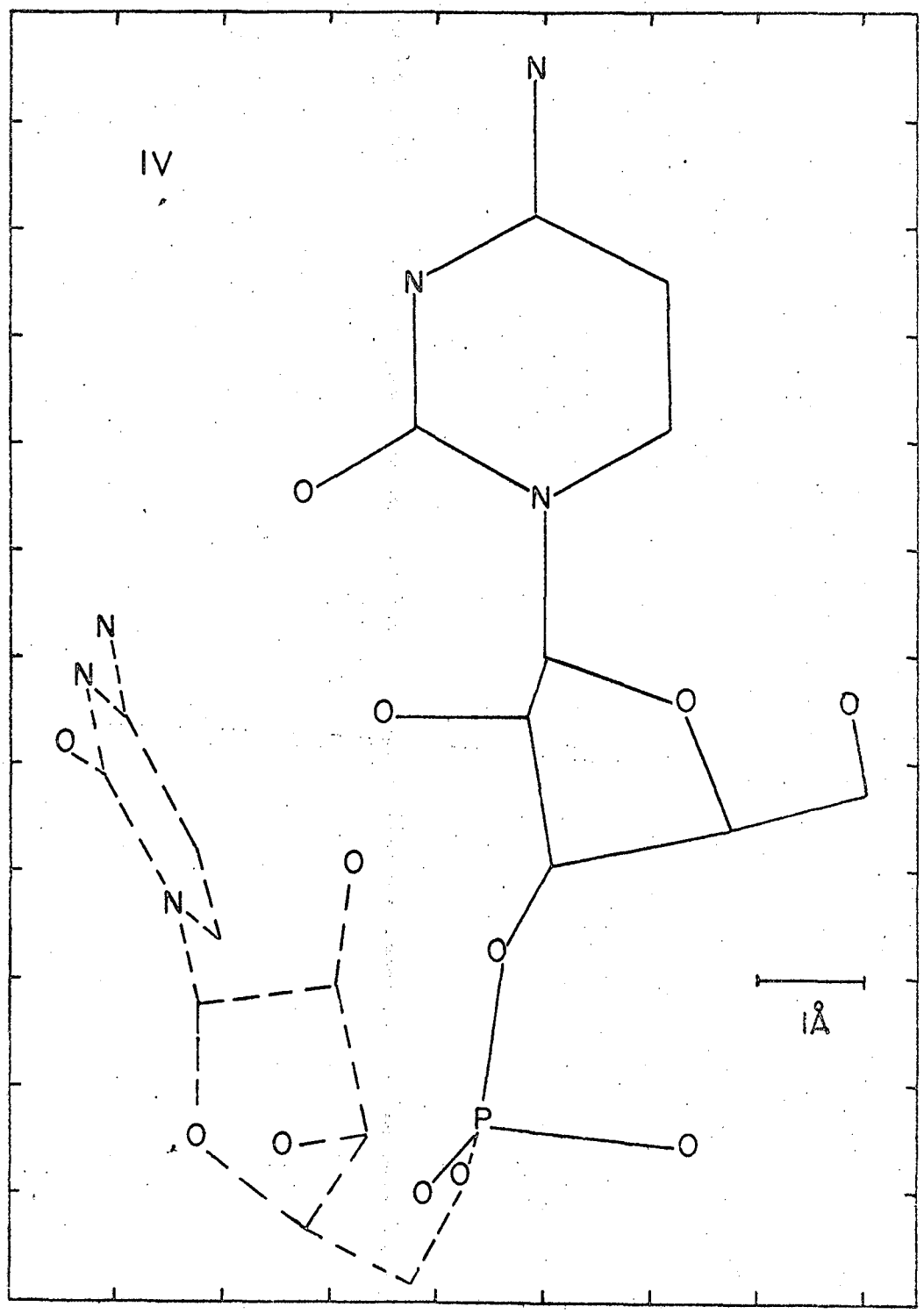
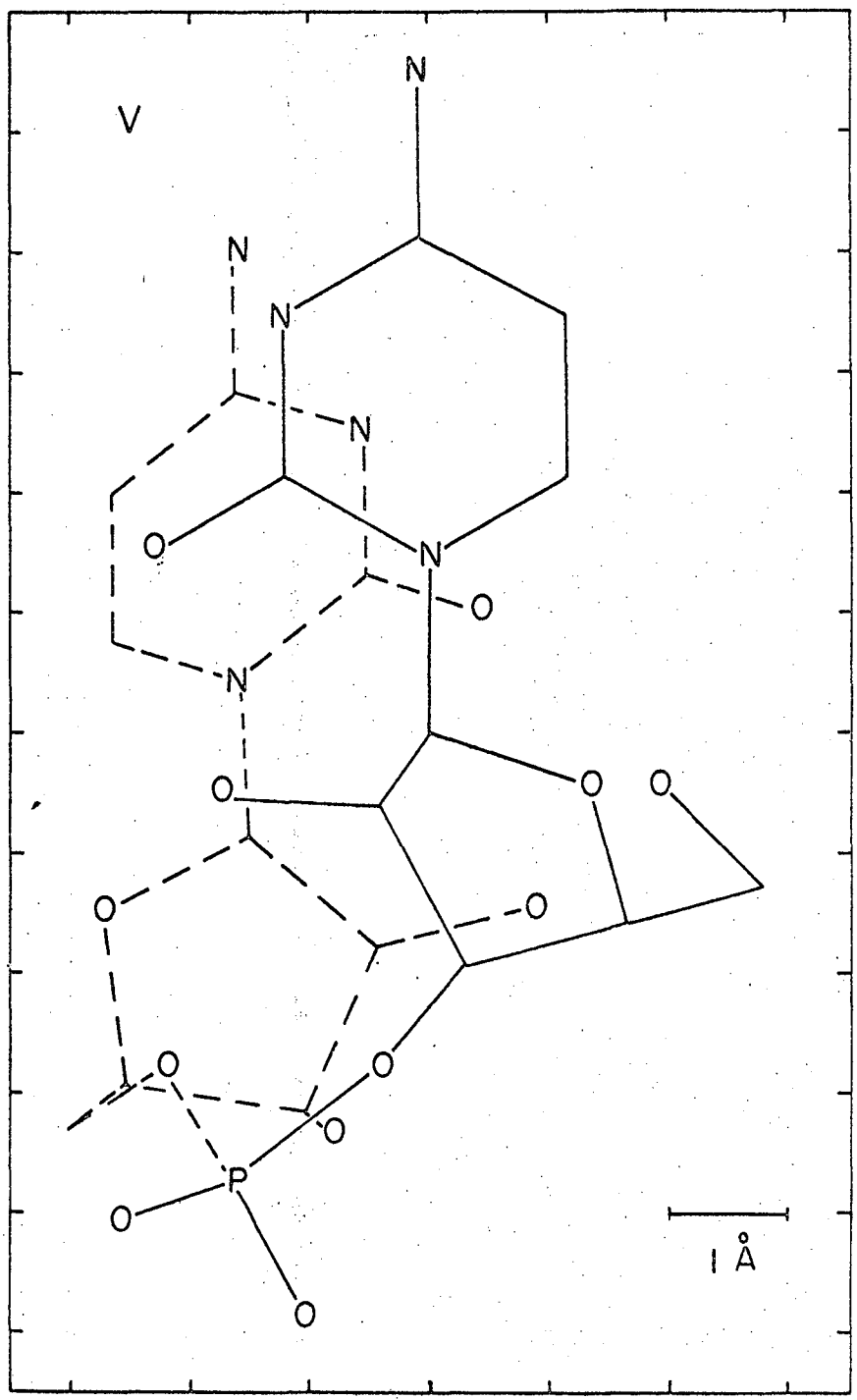



Figure 51 The lowest energy Group III conformation. The 5' nucleoside is indicated by dashed lines.



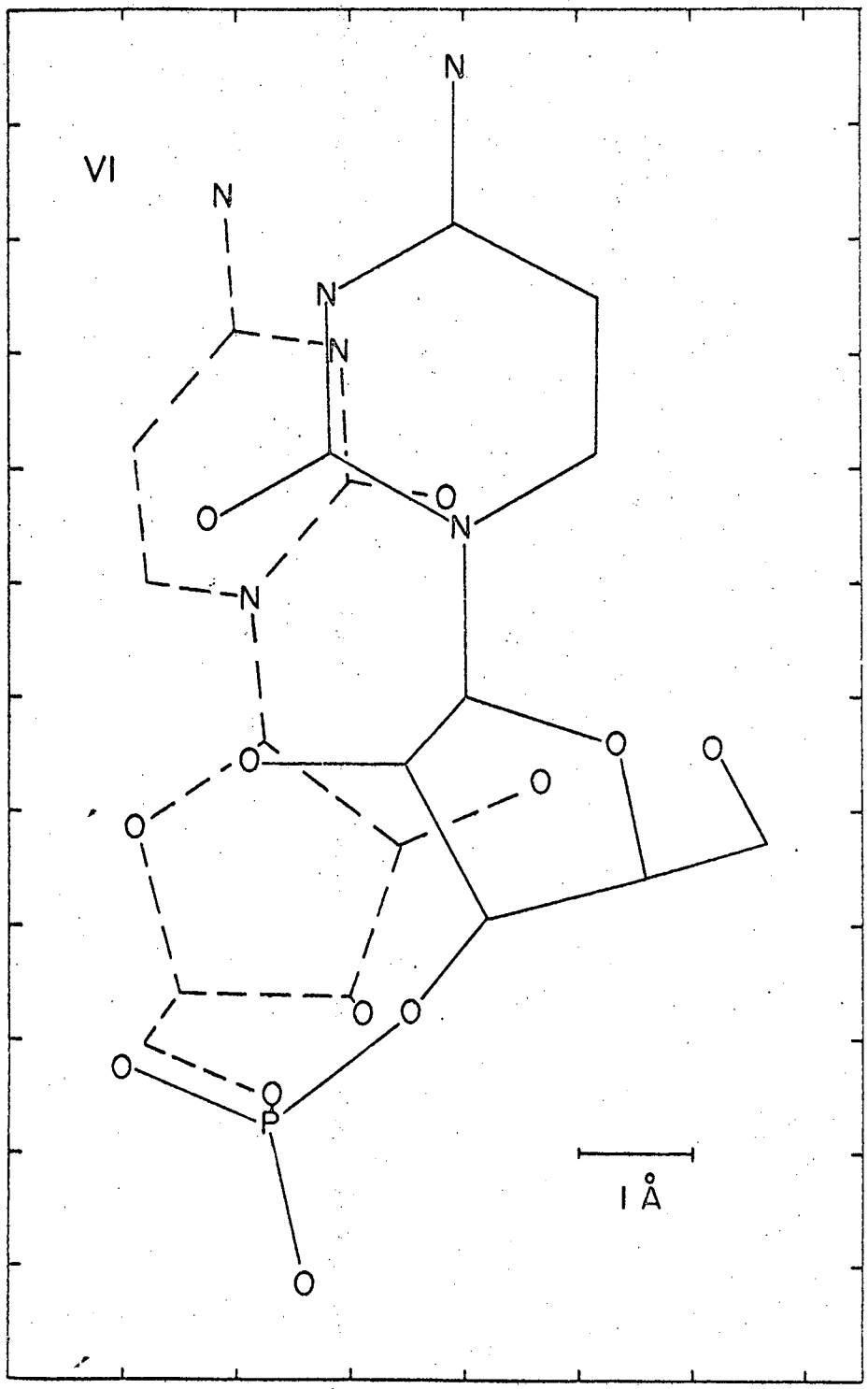

XBL6711-5748

Figure 52 The lowest energy Group IV conformation. The 5' nucleoside is indicated by dashed lines.



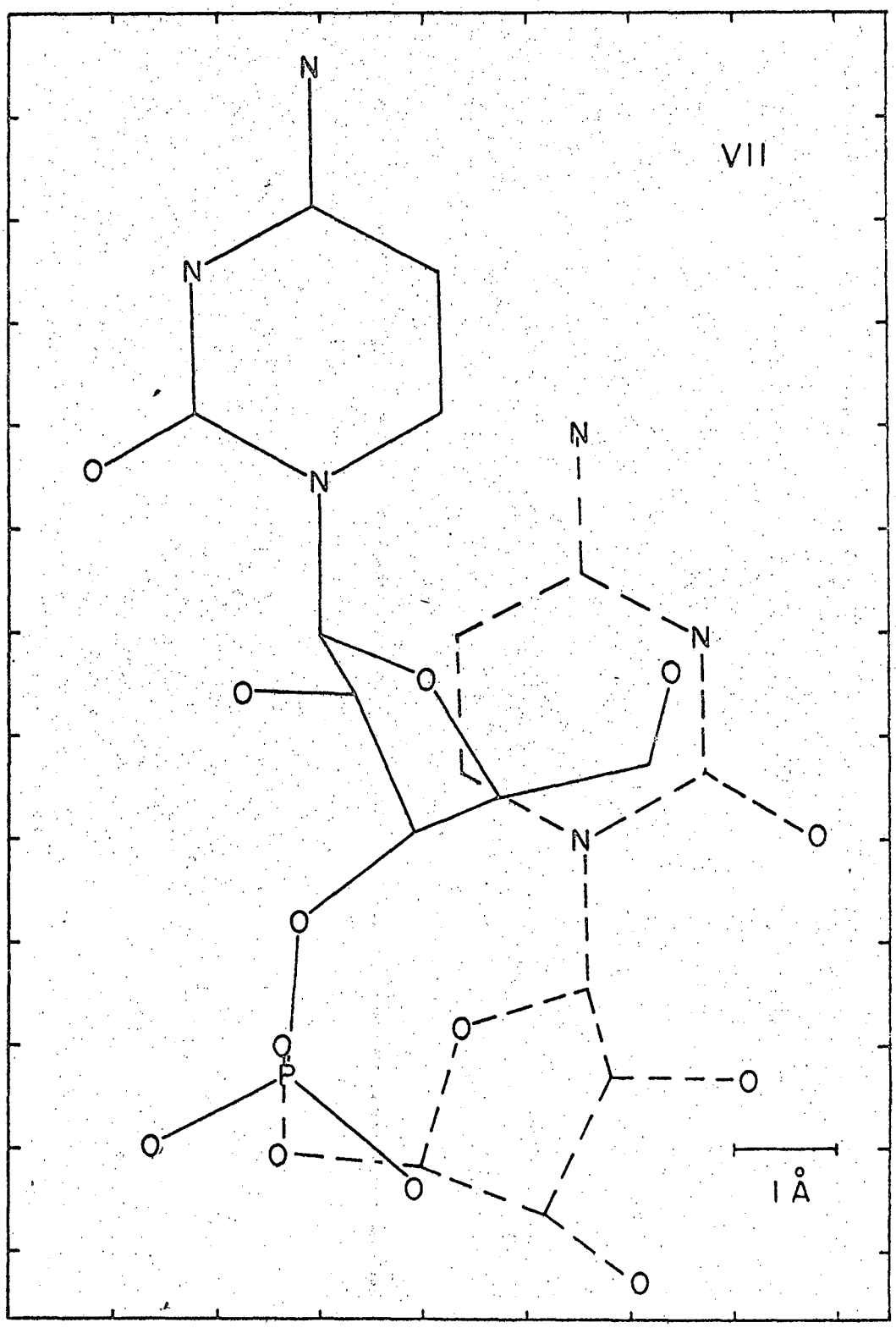

XBL6711-5749

Figure 53 The lowest energy Group V conformation. The 5' nucleoside is indicated by dashed lines.




XBL6711-5750

Figure 54 The lowest energy Group VI conformation. The 5' nucleoside is indicated by dashed lines.




XBL6711-5751

Figure 55 The lowest energy Group VII conformation. The 5' nucleoside is indicated by dashed lines.

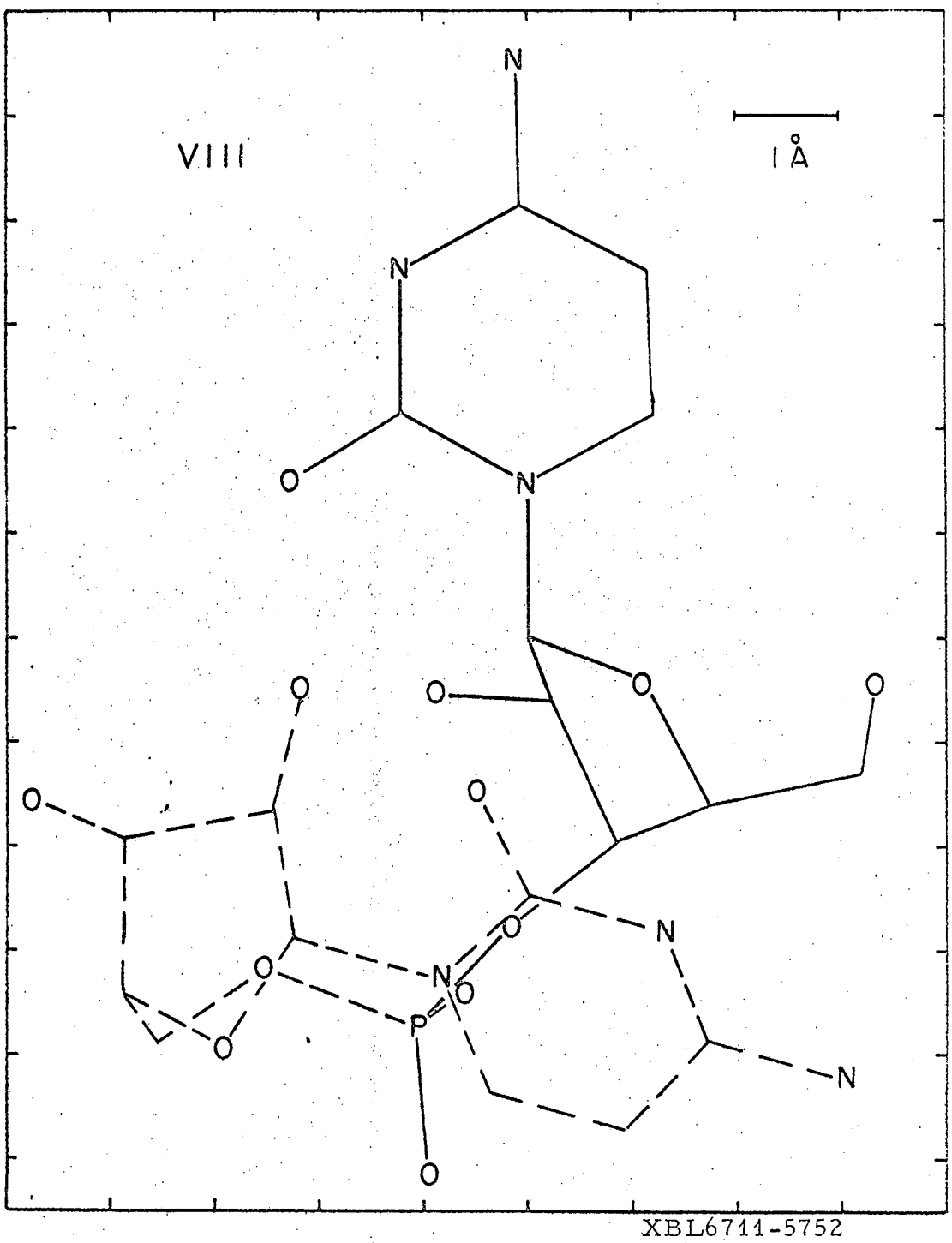



Figure 56 The lowest energy Group VIII conformation. The 5' nucleoside is indicated by dashed lines.

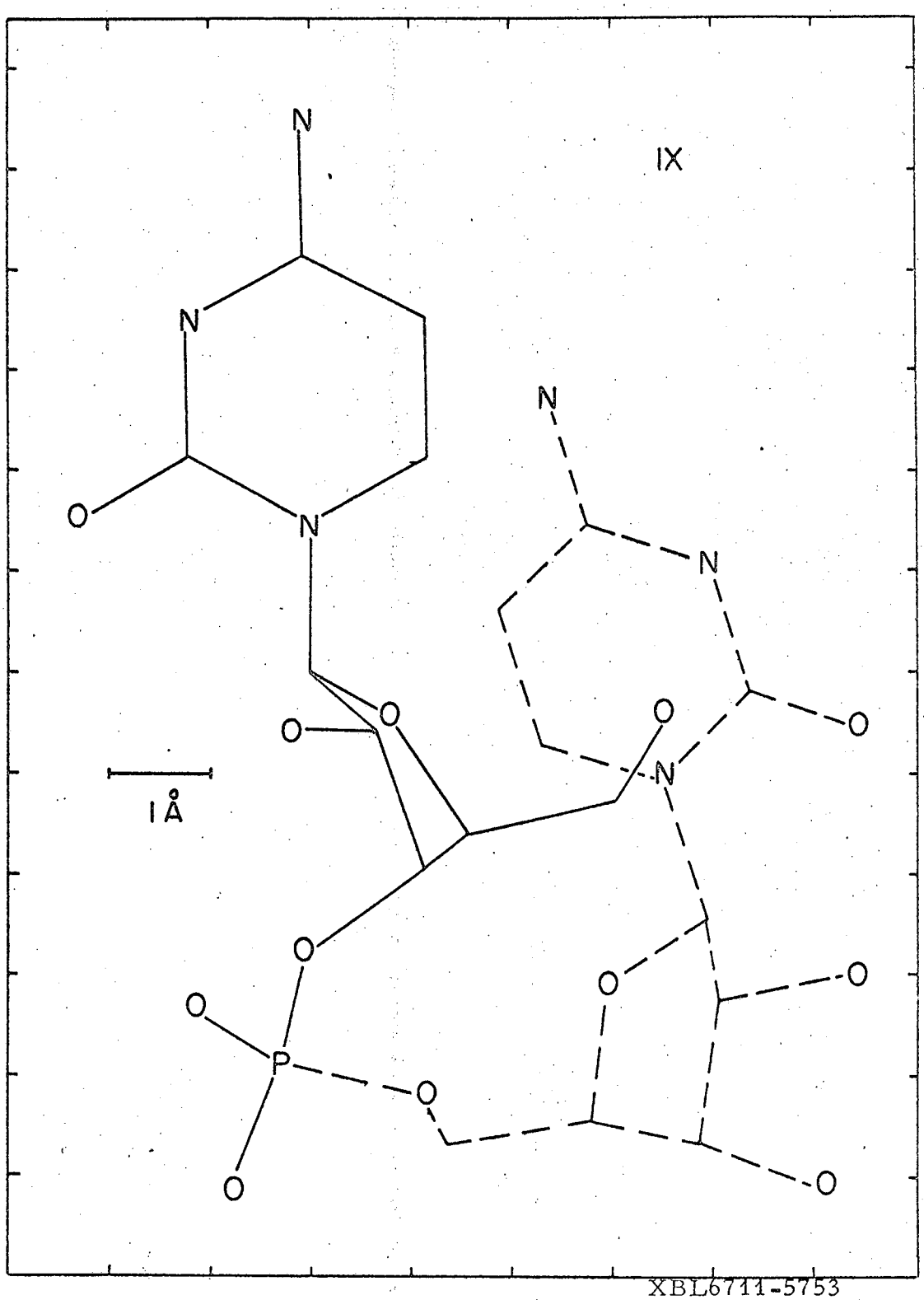
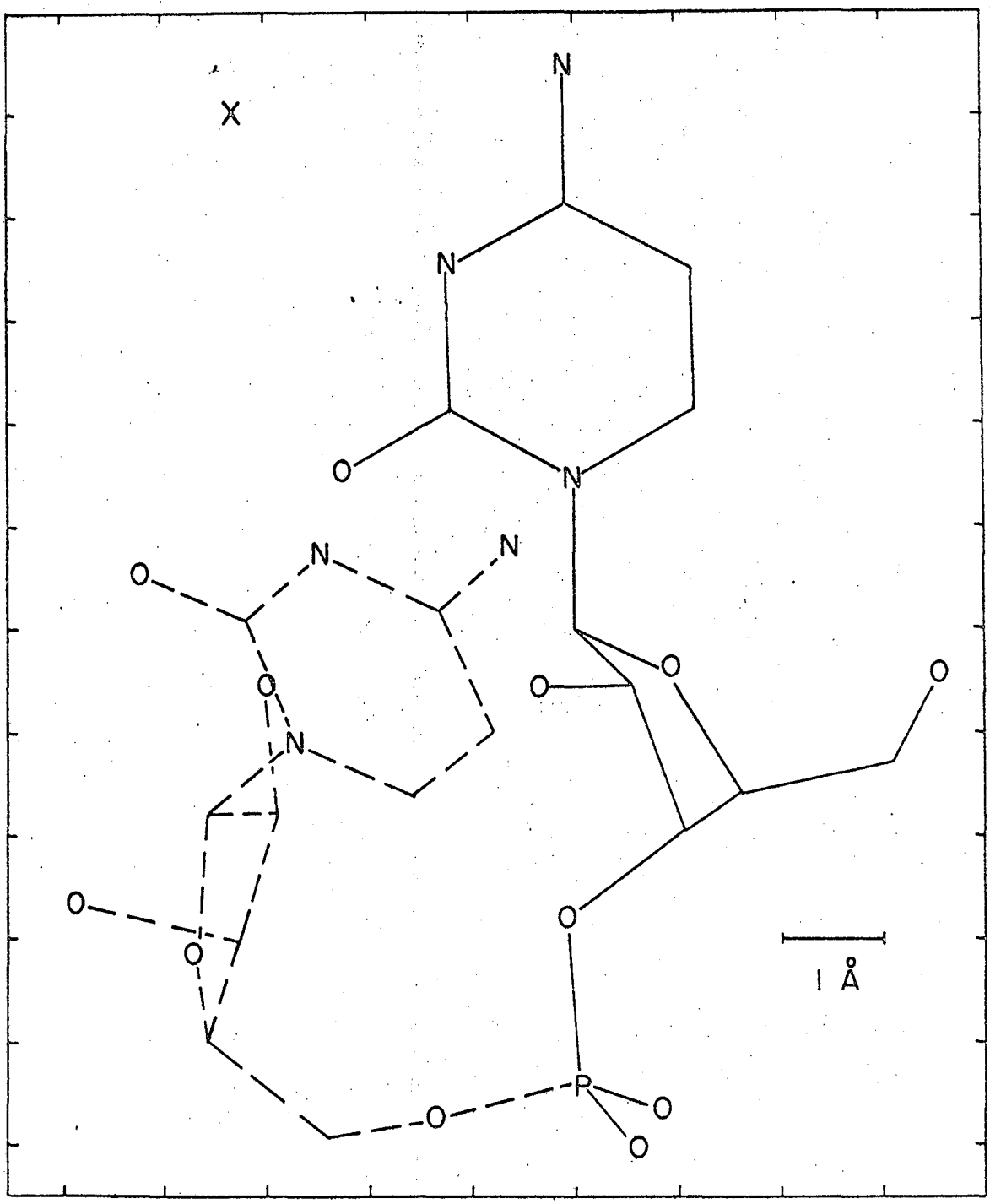



Figure 57 The lowest energy Group IX conformation. The 5' nucleoside is indicated by dashed lines.




XBL6711-5754

Figure 58 The lowest energy Group X conformation. The 5' nucleoside is indicated by dashed lines.

the first part of this thesis.)

Summarizing these descriptions, 21 of the 22 lowest energy conformations of CpC have the two cytosines close enough to one another so that the solvent can not completely isolate them. The energy of the single "unstacked" conformation is 9 kcal per mole higher than the lowest "stacked" conformation. This compares well with the values of $\Delta H^{o}$ ("unstacking") which have been determined from various physical measurements. The "stacked" conformations vary considerably amongst themselves, with the 5' cytosine found on either side of the 3' cytosine. A number of these structures indicate strong interactions between the 5' cytosine and the 3' ribose. None of these conformations corresponds to structures found in polynucleotide fibers.

Table 23 lists the torsional angles which we have calculated to be energetically most favorable (i.e. those found for the 22 low energy conformations of CpC), those "allowed" by Sasisekharan et al.[166] on the basis of hard core criteria and those found in X-ray diffraction studies of polynucleotide fibers[62,63,104,122,178]. The agreement between our results and those of X-ray studies are generally quite good. In most cases we predict a wider range of low energy torsional angles than are found in the double-strand species. This is to be expected, as the polynucleotide fiber structures represent only one special type of oligonucleotide conformation. The noteable deviations of the two sets of results are: the glycosidic bond, where the

TABLE 23

TORSIONAL ANGLES FOUND IN NUCLEIC ACID STRUCTURES

| Bond | These Calculations (22 lowest energy conformations of CpC) | Hard Core Repulsions[166] | Polynucleotide Fibers 62,63,104,122,161,178 |
|---|---|---|---|
| Glycosidic | -60 to -5 | not calculated | -90 to -70<br>-15 to +5 |
| $C_2'-O_2'$ | 5 to 40<br>250 to 340 | | |
| $C_3'-O_3'$ | 200 to 290 | 260 to 270 | ~147<br>200 to 225<br>~241 |
| $O_3'-P$ | ~5<br>100 to 135<br>200 to 320 | not calculated | ~212<br>275 to 295 |
| $P-O_5'$ | 100 to 135<br>160 to 175<br>200 to 300 | not calculated | 280 to 290<br>~315<br>~341 |
| $O_5'-C_5'$ | 105 to 120<br>160 to 175<br>235 to 250 | 150 to 210 | 130 to 145<br>160 to 170 |
| $C_5'-C_4'$ | ~240<br>280 to 325 | 50 to 80<br>140 to 200<br>280 to 310 | 285 to 315 |

B and C forms of DNA deviate unexplainably from both our predictions and the results of crystal studies of small molecules, the $C_3'$-$O_3'$ bond, where the B form of DNA has a torsional angle which we consider to be strongly forbidden in RNA structures because of repulsive interactions involving $O_2'$, and the $O_5'$-$C_5'$ bond, where the barrier to rotation is reasonably low (See Figure 40) and our results may be limited by the small number of conformations we have considered. The agreement of the torsional angles allowed on the basis of hard core considerations and those found in polynucleotide fibers is not as good. In the case of the $C_3'$-$O_3'$ bond the two regions are mutually exclusive.

Table 23 illustrates the point that the energetically favorable torsional regions vary in size from one bond to another. There is considerable variation in the torsional angles of the $O_3'$-P and P-$O_5'$ bonds among the conformations of CpC while there is little variation in torsional angles for the $C_4'$-$C_5'$ bond. The latter is in contrast to the calculations of near-neighbor interactions in potentials for rotations about internucleoside bonds. (It is a result of interactions of atoms which are not close to one another in the covalent bonding scheme.) This information will be of great help if this type of calculation is to be extended to larger molecules. We may be reasonably confident of not missing any low energy conformations if we limit the calculation to torsional angles of 280° to 325° for rotations about the $C_4'$-$C_5'$ bond. Almost all

orientations will have to be surveyed for the P-$O_5$' bond.

An interesting feature found in these calculations is the change of $\phi_{CN}$ in going from nucleosides to dinucleoside phosphates. The values of $\phi_{CN}$ for either the 3' or 5' linked nucleoside in CpC are approximately -10° or -15°, about 20° more positive than the value found in C. Since this occurs in both nucleosides we can say that it is not simply caused by a change in the orientation of the 2' OH. (In the 3' linked nucleoside the 2' OH is allowed to seek its lowest energy orientation while in the 5' nucleoside it is held in the position found in crystals of Cp.) This effect is caused by interactions between the nucleosides. The nature of these interactions is not obvious.

This has implications concerning the origins of the ORD in dinucleoside phosphates. In the case of CpC, the ORD of the component nucleosides is approximately as large as the ORD which is caused (either directly or indirectly) by interaction of the nucleosides[211]. The two ORD's also have the same qualitative shape. Since we expect the ORD of nucleosides to be dependent on $\phi_{CN}$, the contribution to the ORD caused by a 20° change in $\phi_{CN}$ may be important and should be considered when comparing the ORD calculated for dinucleoside phosphates with the experimental ORD.

The results of these calculations seem reasonable when considered one torsional angle at a time. Our calculations of $\phi_{CN}$ generally agree with the results of

both crystal and solution studies. Our predictions for the most stable torsional angles seem to agree well with those found in polynucleotide fibers. Why then do we not calculate right-handed "stacked" conformations of CpC to be energetically favorable? Experimental studies indicate that CpC probably does exist in this structure in the low temperature limit. There are three possible causes for this discrepancy which occur to us at this time: those associated with the execution of the calculation, the choice of geometries and the choice of potentials.

We may have introduced errors into our calculation by using a 40° grid to search for the low energy conformations. Because this grid is coarse, there is a chance that no point on this grid, when minimized, would lead to the true low energy conformation. There is also the possibility that we missed low energy conformations by choosing only the lowest 25 points on the grid for minimization.

In addition, our choices of geometry were to a large part, arbitrary. The computations of CpC were so lengthy that we did not try alternative geometries. Modest changes in the furanose structure (i.e. puckering) may lead to different conformations. On the other hand, it does not seem likely that small differences in the base of phosphate geometries would affect our calculations appreciably.

We feel that the most likely cause for this discrepancy comes from the choice of potential. In particular, the absence of any solvent effects is probably a major source of difficulty. The most unexpected result of our calculations

was the strong degree of base-ribose interaction. Ribose is a very soluble, very hydrophilic molecule. The sugars will be strongly hydrated, greatly reducing interaction of the ribose with other groups (e.g. bases). On the other hand, the bases are not as hydrophilic and will not be solvated as strongly. These two effects should result in the water reducing base-sugar interactions more than base-base interactions, tending to stabilize structures which are suggested by experimental results (i.e. those with strong base-base interactions).

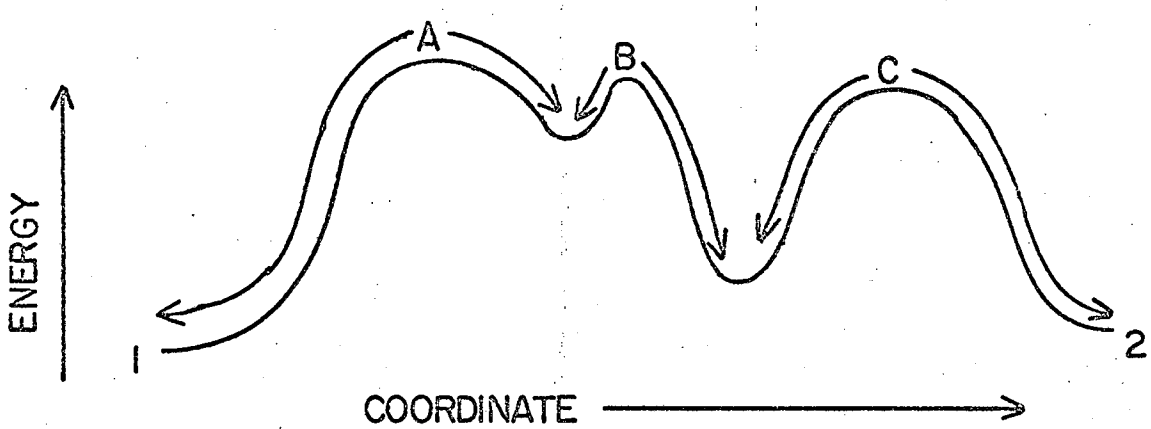
Solvent effects might be introduced into calculations by reducing the charges and polarizabilities of the ribose atoms to approximate the damping effect of the solvent. An alternative method would be to introduce a set of dielectric constants, with higher dielectric constants for interactions involving ribose atoms (as opposed to base-base interactions). We make no claims for the advantages of these methods over other possible ones. Since the approach to the problem is an empirical one, approximations must be tested before a best approach can be chosen.

PATHS BETWEEN CONFORMATIONS

It is important to know whether some conformations are separated from others by essentially impenetrable high energy barriers. To see whether this was the case, we found a path between two very different conformations. (Numbers 1 and 4, Table 22) In order to get from one conformation to the other, one cytosine must be brought

around the other base. This movement is as radical a change as we can imagine. If high energy barriers are to be encountered, we would expect them in this case.

We searched for the path by considering a straight line between the two conformations in the eight dimensional configuration space. We chose seven points, approximately equally spaced and at the nearest 5° interval from the line. Starting at each of these points the minimization routine, FINDMIN, was used to find a close local minimum. This gave us a number of short paths in the region between the two conformations, all leading to local minima. Considering the problem in just one dimension, the situation can be pictured in the following way:




The arrows indicate the paths of the minimizations. They specify the energy over the complete path from state 1 to state 2 with the exceptions of small regions A, B and C. Because our problem is in eight dimensions, the choice of path determines the barriers which must be crossed. When the straight line path seemed to be going over very

high energy regions we tried to circumvent them by trying slightly different paths as indicated by these or previous minimization results. This can be illustrated by considering the changes in torsional angles for the $C_2'$-$O_2'$-, $C_3'$-$O_3'$ and $O_3'$-P bonds in the starting and final conformations. The torsional angles for $C_2'$-$O_2'$ and $C_3'$-$O_3'$ are the same at the beginning and end of the path, 335° and 290° respectively. The torsional angle for the $O_3'$-P bond changes from 125° to 265°. Considering contributions of near neighbor interactions to the potential for rotating about the $C_3'$-$O_3'$ and $O_3'$-P bonds (Figure 41) we see that a path going from (290, 125) directly to (290, 265) goes through a region of reasonably high energy. It would seem much more advantageous to take a slightly more circuitous route, with the torsional angle for the $C_3'$-$O_3'$ bond decreasing slightly near the middle of the path. Not only does this happen, but during the minimization process the 2' OH group swings out of the way (The torsional angle for the $C_2'$-$O_2'$ bond goes from 335° to 35°.) which has the effect of moving the edge of the high energy region to larger values of the torsional angle for the $C_3'$-$O_3'$ bond.

The lowest energy path we were able to find for bringing the 5' cytosine from one side of the 3' cytosine to the other is outlined in Table 24. The arrows indicate the paths of direct minimization (starting at the tail and ending at the head). There are five gaps in the path that are unaccounted for. However they are small gaps. We do not expect them to add appreciably to the heights of saddle points.

TABLE 24

PATH FROM GROUP I TO GROUP II

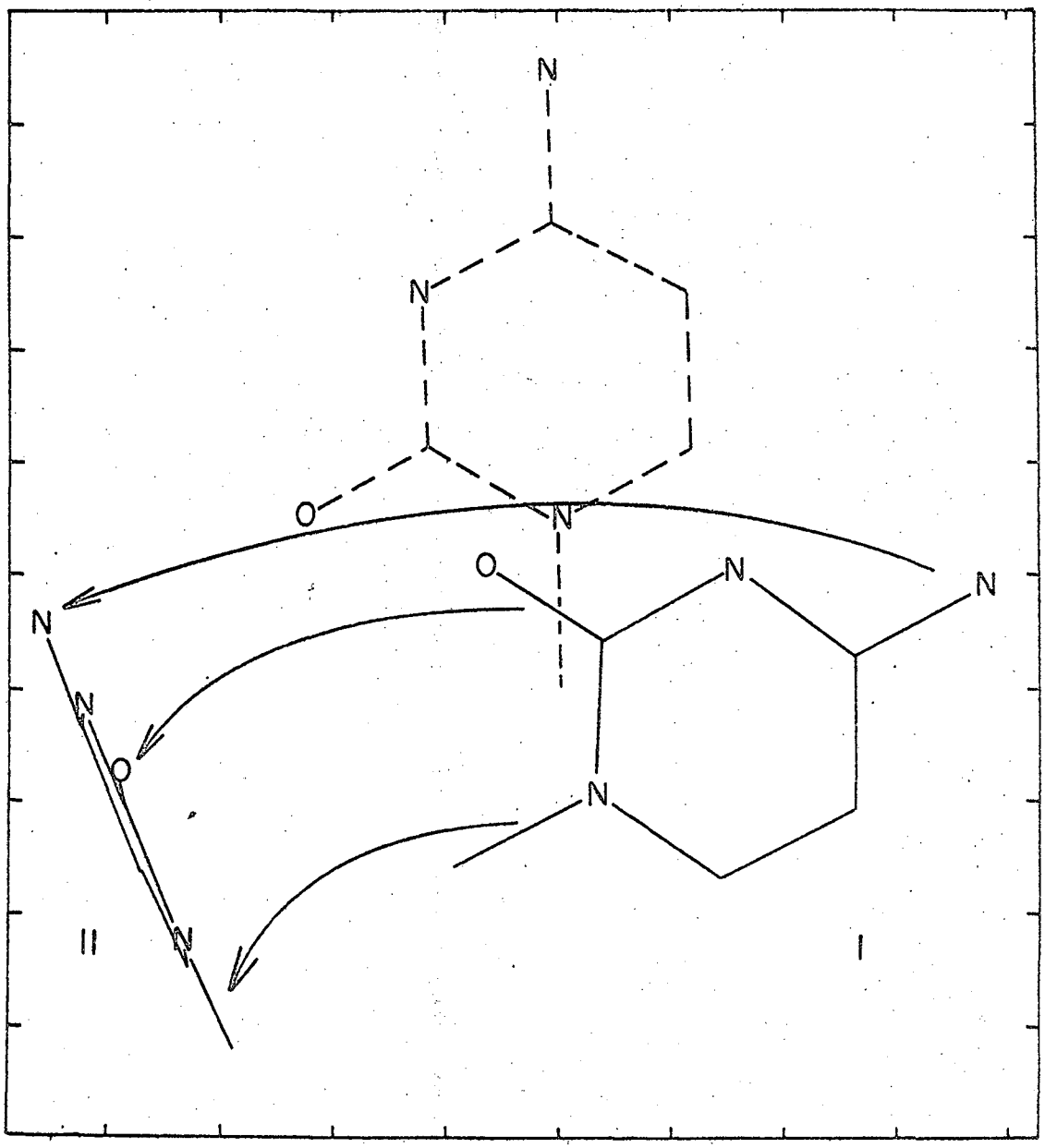
| TORSIONAL ANGLES | | | | | | | | ENERGY |
|---|---|---|---|---|---|---|---|---|
| 3'GLYCOSIDIC | $C_2'-O_2'$ | $C_3'-O_3'$ | $O_3'-P$ | $P-O_5'$ | $O_5'-C_5'$ | $C_5'-C_4'$ | 5'GLYCOSIDIC | (KCAL/MOLE) |
| -10 | 335 | 290 | 120 | 100 | 245 | 320 | -15 | -134.3 |
| ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | |
| -20 | 340 | 290 | 125 | 110 | 245 | 320 | -20 | -129.0 |
| -20 | 335 | 285 | 135 | 115 | 250 | 320 | -25 | -128.2 |
| ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | |
| -15 | 330 | 285 | 155 | 125 | 250 | 320 | -15 | -119.1 |
| -15 | 335 | 280 | 160 | 130 | 250 | 320 | -20 | -120.9 |
| ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | |
| -25 | 30 | 280 | 220 | 125 | 255 | 320 | -35 | -134.1 |
| -25 | 35 | 275 | 230 | 130 | 250 | 320 | -35 | -132.4 |
| ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | |
| -15 | 340 | 285 | 240 | 140 | 250 | 320 | -15 | -121.5 |
| -15 | 340 | 290 | 245 | 150 | 255 | 320 | -15 | -123.1 |
| -15 | 340 | 285 | 255 | 155 | 250 | 320 | -10 | -122.5 |
| ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | |
| -15 | 335 | 290 | 265 | 160 | 255 | 325 | - 5 | -133.2 |

The path from conformations of the group I type to those of the group III type turns out to be one which can be visualized easily. First, the 5' cytosine slides over the 3' cytosine until they are no longer overlapping (group II). (Figures 59 and 60) Then the 5' cytosine is pivoted (with the $C_5$-$C_6$ region moving less than the carbonyl region) until it is tucked under the 3' cytosine with the orientation of the two cytosines opposite each other (group III). (See Figure 51.) There are barriers of approximately equal height (~15 and ~13 kcal per mole) for the sliding motion and the pivoting motion.

Using the Arrhenius unimolecular rate expression (with an approximate frequency factor[8]) we can estimate the rate of exchange between these two conformations.

$$K = \frac{kT}{h} e^{-\Delta H/kT}$$

Here K is the unimolecular rate constant; k is Boltzmann's constant; h is Planck's constant; T is the absolute temperature; $\Delta H$ is the energy of the highest saddle point which must be crossed in order to get from one state to another (15 kcal per mole). This gives us a value of $K = 60\ sec^{-1}$. A reasonable time would be very much shorter than the unwinding times of short DNA segments (.5 sec for $MW \sim 10^6$ [39]), making this the calculated relaxation time considerably longer than expected.




XBL6711-5755

Figure 59 The movement of the 5' cytosine from Group I to Group II.

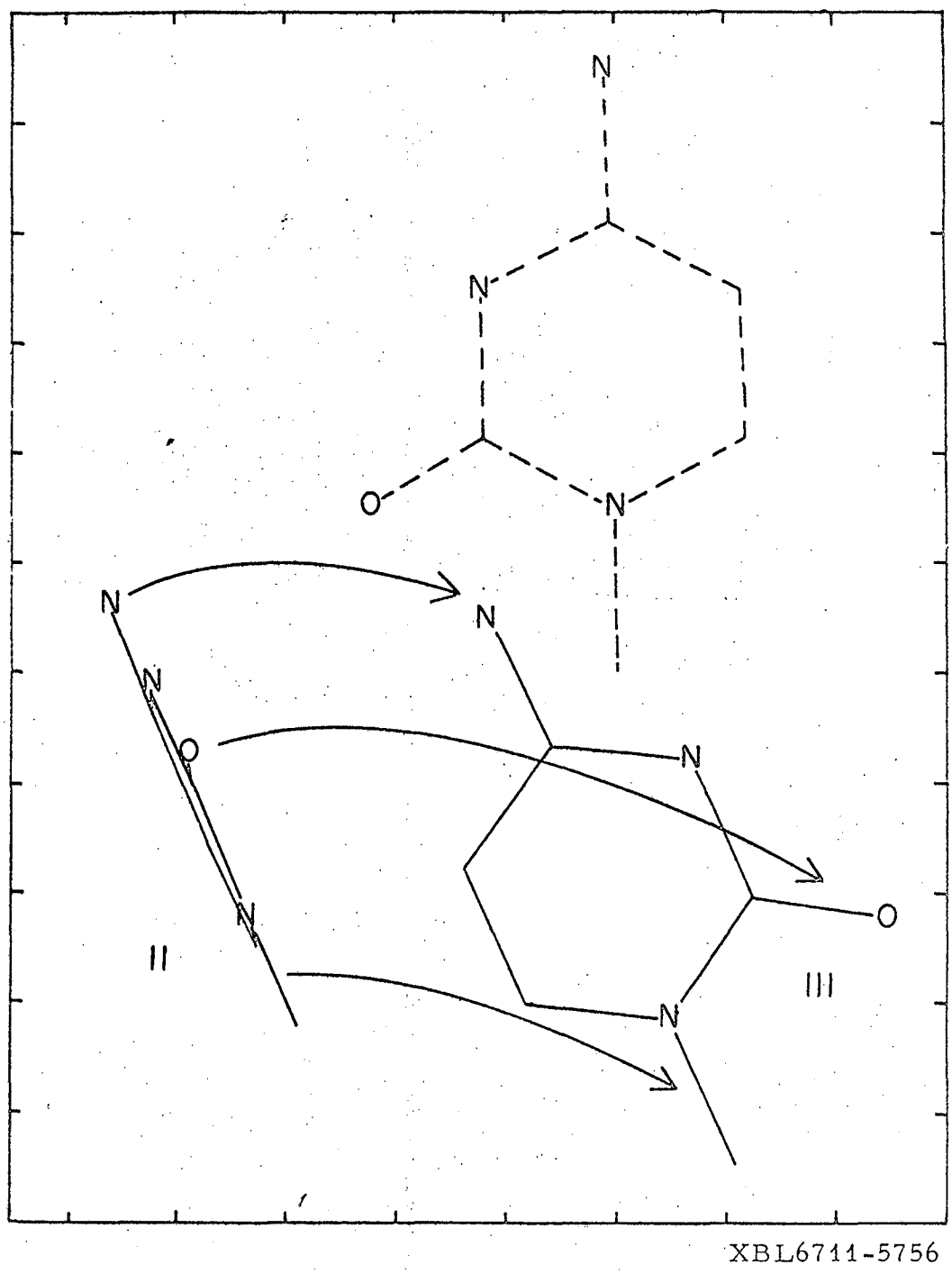



Figure 60 The movement of the 5' cytosine from Group II to Group III.

There are two reasons for saying that the actual rate will be faster. First, there is no reason why there can not be a very different path between the two conformations having lower saddle points. If a more favorable path exists, then the overall rate constant will be approximately the rate constant for that path, rather than the one already discovered. The second reason for considering the rate of exchange as being the slow limit comes from considerations of possible distortions of the presumably "rigid" groups. Large energy gradients (usually associated with high energies) will tend to deform the normally "rigid" groups. The resulting deformations will have the general effect of reducing energy differences between high and low energy conformations. That means that the heights of the saddle points (which must be traversed in going from the starting to final conformations) are lowered. Calculations on dipeptides have shown that these effects may be quite pronounced[64]. Unfortunately it is not possible now to quantitatively estimate how much lower the true saddle points are. We expect the exchange rates between these two very different conformations to be considerably faster than 60 $sec^{-1}$, probably rapid enough that we do not have to consider the possibilities that some conformation of CpC is not attainable for kinetic reasons.

One added benefit we got from the calculations of this path was the discovery of three new low energy conformations. Originally we thought that the conformation described by the vector (-15, 335, 280, 160, 245, 320, -15) of torsional

angles was the lowest energy conformation in group III. (See Table 22) We also discovered two new low energy conformations in group II. This illustrates the fact that we have no guarantee that we have found all conformations corresponding to low energy minima.

ENTROPY

All of our work up to this time has been directed towards calculating the energies of conformations of nucleic acid fragments. We have neglected contributions of entropy. It is the free energy, F = H-TS which determines the population distribution among different conformations at constant temperature and pressure. If we consider a conformation to be the rigid structure described by the set of torsional angles then there will be no entropy term. (All conformations have unit degeneracy.) However, if we define a conformation as the region in configuration space which is easily accessible to the rigid structure by small, thermal oscillations (which is what the experimentalist studies), then an entropy term must be added to the energy.

Theoretical studies have shown that the contribution to the entropy by internal rotations is approximated well by the classical expression, provided the moments of inertia are large[143]. Since this is usually the case with dinucleoside phosphates we have used the equation:

$$S = R \ln \int e^{-E/RT} \, d\tau$$

to compute the entropy differences between the different types of conformations which we have found in these calculations. In this equation R is the gas constant, E is the energy of the conformation, T is the absolute temperature and the integral is performed over phase space. Since the energy, E, depends only on the geometry of the compound (rather than the momentum also) we need integrate only over configuration space, rather than all of phase space.. Unfortunately ours is an 8 dimensional space. Multiple integrations of this type are very time consuming. If we were to do the integration numerically using energies for 5 orientations of each bond near the equilibrium position (which would only give us a hint as to what the answer might be) it would require $5^8$ new energy calculations and the very laborious 8 dimensional integration. This corresponds to hours of computing time. Because the large investment might not yield any worthwhile results (i.e. we would have no reliable estimate of our error) we tried to estimate the entropies in a different way.

If the power series describing the energy surface was additive in terms of functions of the individual torsional angles, $\phi_1$, $\phi_2 \ldots \phi_8$ (i.e. $V=V(\phi_1)+V(\phi_2)+\ldots V(\phi_8)$, then the integral for the entropy would be reduced to the product of 8 single integrals for each conformation. Even if these integrals had to be evaluated numerically the task did not seem too difficult. The assumption of the additivity of the potentials is equivalent to assuming that partial derivatives of more than one variable (e.g.

$\partial^2 V/\partial\phi_i \partial\phi_j$) are zero. Preliminary calculations showed that terms of the type $\partial^2 V/\partial\phi_i \partial\phi_j$ were frequently larger than $\partial^2 V/\partial\phi_i^2$, indicating that the approximation of additive potentials could not be used without introducing considerable errors.

We felt that the best we could do now would be to make a very crude estimate of the variation of the entropies among low energy conformations. This was done by assuming the potential as additive and quadratic in form (i.e. $V = \sum_{i=1}^{8} a_i(\phi_i - \phi_{io})^2$). This is a bad assumption, but all we expect to estimate is how much entropies may vary from one conformation to another. The coefficient of the quadratic term is defined by:

$$a_i = \frac{1}{2} \frac{\partial^2 V}{\partial\phi_i^2}$$

These derivatives are easy to calculate analytically. The key step is the use of the chain rule:

$$\frac{\partial V}{\partial\phi_i} = \sum_a \sum_b \frac{\partial r_{ab}}{\partial\phi_i} \frac{\partial V}{\partial r_{ab}}$$

where the double sum is overall pairwise interactions. Since V is an explicit function of the interatomic distances, $r_{ab}$, the calculation of $\frac{\partial V}{\partial r_{ab}}$ is trivial. $\frac{\partial r_{ab}}{\partial\phi_i}$ is calculated by converting from cartesian to cylindrical coordinates with the z axis along the $i^{th}$ bond, with $r_{ab}$

now a function of the cylindrical coordinates of atoms a and b. Since $\phi$ (usually denoted as the cylindrical coordinate $\theta$ ) is the only cylindrical coordinate to change when rotation takes place about the bond, the derivative is easy to compute. The second derivative is determined by applying the chain rule again. The second partial derivatives for rotations about the 8 bonds for the lowest energy conformations (see Table 22) are listed in Table 25. The size of the second derivative does not vary much among the members of each group of conformations. Table 25 also lists average values of the second derivatives for groups I, II, III and IV.

The expression for the entropy of a conformation:

$$S = \sum_{i=1}^{8} S_i = \sum_{i=1}^{8} R \ln \int e^{-E_i/RT} \, d\phi_i$$

becomes:

$$S = \sum_{i=1}^{8} R \ln \int e^{-\frac{a_i \phi_i^2}{2RT}} \, d\phi_i$$

when the quadratic potential is inserted. All that remains now is the choice of the limits of integration. Since the exponential function decreases very rapidly we do not introduce any appreciable error by integrating from $-\infty$ to $\infty$. (Normally the integrals would be done from $-180°$ to $+180°$.) This integral of the error function is well known and is proportional to $\frac{1}{\sqrt{a_i}}$ . Thus, the expression

TABLE 25 - $\frac{\partial^2 V}{\partial \phi_i^2}$ (CAL PER MOLE DEG$^2$)

| CONFORMATION (SEE TABLE 22) | GROUP | 3'GLYCOSIDIC | $C_2'-O_2'$ | $C_3'-O_3'$ | $O_3'-P$ | $P-O_5'$ | $O_5'-C_5'$ | $C_5'-C_4'$ | 5'GLYCOSIDIC |
|---|---|---|---|---|---|---|---|---|---|
| 1 | I | 43 | 8 | 160 | 219 | 455 | 666 | 1070 | 57 |
| 2 | II | 33 | 4 | 132 | 41 | 146 | 631 | 111 | 37 |
| 3 | I | 31 | 7 | 208 | 158 | 155 | 666 | 746 | 48 |
| 4 | III | 13 | 3 | 250 | 536 | 170 | 802 | 653 | 143 |
| 5 | I | 49 | 6 | 180 | 235 | 199 | 644 | 1456 | 70 |
| 6 | II | 33 | 4 | 97 | 48 | 140 | 434 | 730 | 29 |
| 7 | III | 13 | 3 | 240 | 334 | 383 | 334 | 612 | 88 |
| 8 | IV | 12 | 3 | 139 | 420 | 116 | 435 | 605 | 114 |
| 9 | IV | 74 | 6 | 120 | 33 | 1639 | 446 | 2105 | 93 |
| 10 | I | 28 | 7 | 165 | 175 | 293 | 1047 | 1206 | 62 |
| | | | | AVERAGE | | | | | |
| | I | 40 | 7 | 175 | 190 | 200 | 660 | 1100 | 60 |
| | II | 33 | 4 | 115 | 45 | 145 | 500 | 400 | 30 |
| | III | 13 | 3 | 240 | 400 | 170 | 400 | 630 | 115 |
| | IV | 74 | 6 | 120 | 33 | 1640 | 450 | 2100 | 93 |

TABLE 26

RELATIVE ENTROPIES OF GROUPS OF CONFORMATIONS

| | | |
|---|---|---|
| I | 1.0 | cal/deg mole |
| II | 6.0 | " |
| III | 2.6 | " |
| IV | 0 | " |

for the entropy of a conformation becomes:

$$S = -\sum_{i=1}^{8} \frac{R}{2} \ln \frac{\partial^2 V}{\partial \phi_i^2}$$

Using this expression we have estimated the relative entropies of the conformations of groups I, II, III and IV. These entropies are listed in Table 26.

Entropy differences of 5 e.u. may be commonplace in this type of situation. At 300°K this corresponds to a 1.5 kcal per mole contribution to the free energy which is appreciable compared to the energy differences between conformations. (The energy difference between the two lowest energy conformations we have found for CpC is only 0.2 kcal per mole.)

It is unfortunate that more accurate treatments of the entropy are so difficult to carry our. Once a reasonable degree of confidence is developed in our ability to calculate energies of conformations accurately it may be worthwhile to make the investment necessary to determine more precise values of the entropy differences between conformations. Until that time we will have to remain in the unsatisfying situation of knowing that entropy differences between conformations may be appreciable but that they are extremely difficult to calculate.

CONCLUSION

Although the original goal of these calculations was the determination of realistic potential energy surfaces

for dinucleoside phosphates (so that we could develop models to explain the temperature dependent properties which have been measured already) we can not separate this from the attempt to calculate a priori structures of nucleosides and dinucleoside phosphates in evaluating our success. Such an appraisal will indicate some of the weaknesses of our methods, as well as the directions to go in order to improve them.

Our first results indicate the pyrimidine nucleosides should be in the anti conformation, while purine nucleosides may be either syn or anti. Because the purine nucleosides may be in the syn orientation and consequently would have to flip to the anti orientation to form a double-strand nucleic acid structure, differences in the energies of the syn and anti forms (which depend on the sugar) may cause different stabilities of double-strand ribo- and deoxyribopolynucleotide structures. Experimental studies agree with our prediction of the structure of pyrimidine nucleosides. Unfortunately, the question of the structure of purine nucleosides has not yet been resolved.

Calculations of the orientations of the internucleoside bonds ($C_3'$-$O_3'$, $O_3'$-P, P-$O_5'$, $O_5'$-$C_5'$ and $C_5'$-$C_4'$) indicate that the repulsive interactions of $O_2'$, $O_1'$ and the phosphate oxygens are probably important in causing the observed differences between ribopolynucleotides and deoxyribopolynucleotides (both single- and multiple-strand). We found no indications that hydrogen bonds involving the 2'

OH group played an important role in stabilizing structures.

Calculations of the orientations of CpC yielded some unexpected results. It seems likely that we have not found the true energy minimum. Although the two cytosines were close to one another in the low energy conformations there was little base-base interaction, with the bases exposed to the solvent. In many structures there was much overlap between the cytosine of one nucleoside and the ribose of the other. This may be caused by our not including solvent effects in our calculations, effects which would reduce any interactions involving the sugars. In addition preliminary calculations indicated that entropy differences between conformations may be appreciable.

The calculations on CpC did give us much information necessary if we are to improve and extend these studies. For some bonds almost all torsional angles must be considered, while for other bonds the barriers to rotation are high and only a small fraction of the torsional angles need be considered. This will speed future calculations.

There are two ways for getting around the problems caused by the omission of solvent effects. The first is to try to include such effects in our calculations. We have already mentioned reducing the charges and polarizabilities of the ribose or introducing sets of dielectric constants. The other way would be to do calculations on systems in which solvent effects may be smaller. A dinucleoside phosphate with two pyrimidine bases may very well have been a poor first choice. Perhaps a dinucleoside phosphate

with two purine bases (e.g. ApA) where there is a greater opportunity for extensive base-base overlap (which would exclude the solvent) would have been better. Other structures which would tend to be more compact with less area exposed to the solvent might include longer oligonucleotides (e.g. trinucleoside diphosphates) or double-strand or triple-strand complexes of small oligomers.

Perhaps the greatest problem in these calculations is the lack of suitable checks of these preliminary calculations. The only reasonably conclusive check of our results which exists now is the experimental work on the orientation of pyrimidine ribosides. If there were more checks available we could see which of our calculations gave us reasonable answers. Using this as a starting point we could start varying parameters or introducing new terms. We anxiously await the results of such tests which hopefully will demonstrate that we have already made substantial progress towards understanding the structures of single-strand oligonucleotides and will greatly aid us in extending this technique to even more interesting systems.

# ACKNOWLEDGEMENTS

I would like to take this opportunity to mention some of the many people without whose help this thesis would never have materialized.

It is a genuine pleasure to acknowledge the help of Professor Ignacio Tinoco, Jr. His advice, encouragement and interest throughout my stay in Berkeley has helped me immeasureably in seeing this work through. In addition, his laboratory has been the stimulating environment which is conducive to research. Many members of his group have contributed to this work. In particular, thanks go to Dr. Myron Warshaw for gifts of dinucleoside phosphates, Dr. Daniel Glaubiger and Mr. David Lloyd for discussions about the oscillating dimer model and Dr. Charles Cantor and Dr. Stanley Mandeles for helping a novice to learn a few of the tricks of wet biochemistry.

Discussions with Professors Sunney Chan and Paul Ts'o were invaluable in interpreting the NMR data. Suggestions of Dr. Gary Felsenfeld and Professor Robert Rein helped in choosing experiments and controls. Calculations of the conformations of dinucleoside phosphates were facilitated by communications of unpublished results from Dr. Alberte Pullman and Dr. Michael Spencer.

In addition, I would like to thank Dr. Melvin Klein and Mr. Don Phelps of the Laboratory of Chemical Biodynamics for help in recording NMR spectra, Mr. N. J. Bohac, Mr. C. J. Baugh and Mr. W. W. Anderson, Jr. of the Chemistry

Department Machine Shop for help with the onerous task of building and maintaining much of the experimental equipment, Mr. William Dempster of the Lawrence Radiation Laboratory Computing Center for help in writing and debugging computer programs, Mrs. Susan Emberton for patiently digitizing many of the spectra presented and Miss Delores Orsten for typing this thesis and drawing many of the figures.

This work was supported in part by a predoctoral fellowship from the U. S. Public Health Service.

Finally, this list would be glaringly incomplete without mentioning my wife, Fran, whose patience, understanding and encouragement have helped to make this work possible.

This work was also supported in part by the U. S. Atomic Energy Commission.

REFERENCES

1. Th. Ackerman and H. Rüterjans, Ber. Bunsengesellschaft, 68, 850 (1964).

2. C. B. Anfinsen, E. Haber, M. Sela and F. H. White, Jr., Proc. Nat. Acad. Sci., U. S., 47, 1309 (1961).

3. J. Applequist and V. Damle, J. Am. Chem. Soc., 88, 3895 (1966).

4. S. Arnott, F. Hutchinson, M. Spencer, M. H. F. Wilkins, W. Fuller and R. Langridge, Nature, 211, 227 (1966).

5. A. Baglioni, V. M. Ingram and E. Sullivan, Nature, 189, 465 (1961).

6. E. K. Bautz and E. A. Bautz, Proc. Nat. Acad. Sci., U. S., 52, 1476 (1964).

7. G. H. Beaven, E. R. Holiday and E. A. Johnson, The Nucleic Acids, ed. E. Chargaff and J. N. Davidson (New York: Academic Press, 1956), v. I, pp. 493 ff.

8. S. Benson, The Foundations of Chemical Kinetics (New York: McGraw-Hill Co., 1960) p. 250.

9. H. Berthod, C. Giessner-Prettre and A. Pullman, Theoret. Chem. Acta (Berlin) 5, 53 (1966).

10. T. M. Birhstein and O. B. Ptitsyn, Conformations of Macromolecules, trans. S. N. Timasheff and M. J. Timasheff (New York: Interscience Publishers, 1966) pp. 29 ff.

11. D. J. Blears and S. S. Danyluk, J. Am. Chem. Soc., 89, 21 (1967).

12. A. Bondi, J. Phys. Chem., 68, 441 (1964).

13. D. F. Bradley, S. Lifson and B. Honig in Electronic Aspects of Biochemistry, ed. B. Pullman (New York: Academic Press, 1964), p. 77 ff.

14. J. Brahms, J. C. Maurizot and A. M. Michelson, J. Mol. Biol., 25, 465 (1967).

15. J. Brahms, J. C. Maurizot and A. M. Michelson, J. Mol. Biol., 25, 481 (1967).

16. J. Brahms, A. M. Michelson and K. E. Van Holde, J. Mol. Biol., 15, 467 (1966).

17. D. A. Brant and P. J. Flory, J. Am. Chem. Soc., 87, 663 (1965).

18. D. A. Brant and P. J. Flory, J. Am. Chem. Soc., 87, 2791 (1965).

19. D. A. Brant, W. G. Miller and P. J. Flory, J. Mol. Biol., 23, 47 (1967).

20. D. A. Brant and P. R. Schimmel, Proc. Nat. Acad. Sci., U. S., 58, 428 (1967).

21. A. D. Broom, M. P. Schweizer and P. O. P. Ts'o, J. Am. Chem. Soc., 89, 3612 (1967).

22. I. H. Brow, K. B. Freeman and H. E. Johns, J. Mol. Biol., 15, 640 (1966).

23. C. E. Bugg and R. E. Marsh, J. Mol. Biol., 25, 67 (1967).

24. L. G. Bunville, E. P. Geiduschek, M. A. Rawitscher and J. M. Sturtevant, Biopolymers, 3, 213 (1965).

25. C. A. Bush, Thesis, University of California, Berkeley (1965).

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

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

28. C. R. Cantor, personal communication (1966).

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

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

31. M. J. Chamberlin, Federation Proc., 24, 1446 (1965).

32. M. J. Chamberlin and D. L. Patterson, J. Mol. Biol., 12, 410 (1965).

33. S. I. Chan, B. W. Bangerter and H. H. Peter, Proc. Nat. Acad. Sci., U. S., 55, 720 (1966).

34. S. I. Chan and J. H. Nelson, personal communication (1967).

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

36. P. Claverie, B. Pullman and J. Caillet, J. Theoret. Biol., 12, 419 (1966).

37. E. Clementi and D. R. Davis, J. Chem. Phys. 45, 2593 (1966).

38. R. L. Collin, J. Am. Chem. Soc., 88, 3281 (1966).

39. D. M. Crothers, J. Mol. Biol., 9, 712 (1964).

40. D. M. Crothers and B. H. Zimm, J. Mol. Biol., 9, 1 (1964).

41. W. C. Davidon, Argonne National Laboratory Report, ANL 5990 (1959).

42. D. R. Davies, Ann. Rev. of Biochem., 36, 321 (1967).

43. S. Davis, Thesis, University of California, Berkeley (1965).

44. G. Del Re in Electronic Aspects of Biochemistry, ed. B. Pullman (New York: Academic Press, 1964) pp. 221 ff.

45. P. De Santis, E. Giglio, A. M. Liquori and A. Ripamonti, Nature, 206, 456 (1965).

46. H. DeVoe, Nature, 197, 1295 (1963).

47. H. De Voe, J. Chem. Phys., 41, 393 (1964).

48. H. De Voe and I. Tinoco, Jr., J. Mol. Biol., 4, 500 (1962).

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

50. J. Donohue and K. N. Trueblood, J. Mol. Biol., 2, 363 (1960).

51. E. A. Dratz, Thesis, University of California, Berkeley (1966) p. 12.

52. J. T. Edsall, P. J. Flory, J. C. Kendrew, A. M. Liquori, G. Nemethy, G. N. Ramachandran and H. A. Scheraga, Biopolymers, 4, 121 (1966).

53. T. R. Emerson, R. J. Swan and T. L. V. Ulbricht, Biochem. Biophys. Res. Comm., 22, 505 (1966).

54. T. R. Emerson, R. J. Swan and T. L. V. Ulbricht, Biochemistry, 6, 843 (1967).

55. S. W. Englander and J. J. Englander, Proc. Nat. Acad. Sci., U. S., 53, 370 (1965).

56. R. M. Epand and H. A. Scheraga, J. Am. Chem. Soc., 89, 3888 (1967).

57. S. Fuberg, Acta Cryst., 3, 325 (1950).

58. S. Fuberg, C. S. Petersen and S. Rómming, Acta Cryst., 18, 313 (1965).

59. H. Fujioka and H. A. Scheraga, Biochemistry, 4, 2197 (1965).

60. H. Fujoika and H. A. Scheraga, Biochemistry, 4, 2206 (1965).

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

62. W. Fuller, F. Hutchinson, M. Spencer and M. H. F. Wilkins, J. Mol. Biol., 27, 507 (1967).

63. W. Fuller, M. H. F. Wilkins, H. R. Wilson, L. D. Hamilton and S. Arnott, J. Mol. Biol., 12, 60 (1965).

64. K. D. Gibson and H. A. Scheraga, Biopolymers, 4, 709 (1966).

65. K. D. Gibson and H. A. Scheraga, Proc. Nat. Acad. Sci., U. S., 58, 420 (1967).

66. S. J. Gill, M. Downing and G. F. Sheats, Biochemistry, 6, 272 (1967).

67. D. L. Glaubiger, Thesis, University of California, Berkeley (1965).

68. Gmelins Handbuch der Anorganischen Chemie, 8th ed., Lithium Supplementary Volume, System no. 20 (Weinheim: Verlag Chemie, 1960) p. 344.

69. H. B. Gray, Jr., V. A. Bloomfield and J. E. Hearst, J. Chem. Phys., 46, 1493 (1967).

70. R. M. Hamlin, Jr., R. C. Lord and A. Rich, Science, 148, 1734 (1965).

71. G. G. Hammes and H. A. Scheraga, Biochemistry, 5, 3690 (1966).

72. S. Hanlon, Biochem. Biophys. Res. Comm., 23, 861 (1966).

73. D. R. Harris and W. M. Macintyre, Biophys. J., 4, 203 (1964).

74. K. A. Hartman, Jr., J. Phys. Chem., 70, 270 (1966).

75. A. E. V. Haschemeyer and A. Rich, J. Mol. Biol., 27, 369 (1967).

76. A. E. V. Haschemeyer and H. M. Sobell, Acta Cryst., 18, 525 (1965).

77. A. E. V. Haschemeyer and H. M. Sobell, Acta Cryst., 19, 125 (1965).

78. G. R. Haugen and W. H. Melhuish, Trans. Faraday Soc., 60, 386 (1961).

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

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

81. F. B. Howard, J. Frazier, M. N. Lipsett and H. T. Miles, Biochem. Biophys. Res. Comm., 17, 93 (1964).

82. R. M. Huber, Acta Cryst., 10, 129 (1967).

83. J. Iball, C. H. Morgan and H. R. Wilson, Nature, 209, 1230 (1966).

84. M. S. Itzkowitz, personal communication (1967).

85. C. D. Jardetzsky, J. Am. Chem. Soc., 82, 229 (1960).

86. C. D. Jarketzsky, J. Am. Chem. Soc., 84, 62 (1962).

87. C. D. Jardetzsky and O. Jardetzsky, J. Am. Chem. Soc., 82, 222 (1960).

88. L. H. Jensen, personal communication (1967).

89. C. E. Johnson, Jr., and F. A. Bovey, J. Chem. Phys., 29, 1012 (1958).

90. W. C. Johnson, Jr., personal communication (1967).

91. J. Josse, A. D. Kaiser and A. Kornberg, J. Biol. Chem., 236, 864 (1961).

92. M. Karplus, J. Chem. Phys., 30, 11 (1959).

93. G. Kartha, J. Bello and D. Harker, Nature, 213, 862, (1967).

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

95. W. Kauzmann, Adv. in Protein Chem., 14, 1 (1959).

96. J. A. A. Ketelaar, Chemical Constitution (Amsterdam: Elsevier, 1958) p. 91.

97. J. E. Kilpatrick, K. S. Pitzer and R. Spitzer, J. Am. Chem. Soc., 69, 2483 (1967).

98. W. A. Klee and S. H. Mudd, Biochemistry, 6, 988 (1967).

99. J. Kraut and L. H. Jensen, Acta Cryst., 16, 79 (1963).

100. Y. Kyogoku, R. C. Lord and A. Rich., J. Am. Chem. Soc., 89, 496 (1967).

101. Y. Kyogoku, R. C. Lord and A.Rich, Proc. Nat. Acad. Sci., U. S., 57, 250 (1967).

102. J. A. Lake and W. W. Beeman, Science, 156, 1371 (1967).

103. M. E. Lamm and D. M. Neville, Jr., J. Phys. Chem., 69, 3872 (1965).

104. R. Langridge, D. A. Marvin, W. E. Seeds, H. R. Wilson and C. W. Hooper, J. Mol. Biol., 2, 38 (1960).

105. P. D. Lawley, Biochim. Biophys. Acta, 21, 481 (1956).

106. M. Leng and G. Felsenfeld, J. Mol. Biol., 15, 455 (1966).

107. C. Levinthal, Sci. Amer., 214, 42 (1966).

108. C. Levinthal, E. R. Signer and K. Fetherholf, Proc. Nat. Acad. Sci., U. S., 48, 1230 (1962).

109. J. Levshin, J. Exper. Theor. Phys., USSR, 1, 235 (1955).

110. J. Levshin, J. Exper. Theor. Phys., USSR, 1, 244 (1955).

111. L. K. Li, J. R. Riehm and H. A. Scheraga, Biochemistry, 5, 2043 (1966).

112. C. C. Lin and J. D. Swalen, Rev. Mod. Phys., 31, 841 (1959).

113. M. N. Lipsett, J. Biol. Chem., 239, 1256 (1964).

114. M. N. Lipsett, L. A. Heppel and D. F. Bradley, J. Biol. Chem., 236, 857 (1961).

115. A. M. Liquori, J. Polymer Sci., C12, 209 (1966).

116. D. A. Lloyd, personal communication (1967).

117. R. Lumry, R. Biltonen and J. F. Brandta, Biopolymers, 4, 917 (1966).

118. V. Luzzati, A. Mathis, F. Masson and J. Witz, J. Mol. Biol., 10, 28 (1964).

119. H. R. Mahler, B. Kline and B. D. Mehrotra, J. Mol. Biol., 9, 801 (1964).

120. H. Margenau, Rev. Mod. Phys., 11, 1 (1939).

121. A. E. Martell (compiler) Stability Constants, Sec. II, Organic Ligands, Publication No. 17 (London: The Chemical Society, 1964) p. 649.

122. D. A. Marvin, M. Spencer, M. H. F. Wilkins and L. D. Hamilton, J. Mol. Biol., 3, 547 (1961).

123. J. Massoulie, A. M. Michelson and F. Pochon, Biophys. Biochim. Acta., 114, 16 (1966).

124. C. C. McDonald, W. D. Phillips and J. Lazar, J. Am. Chem. Soc., 89, 4224 (1967).

125. E. G. McRae, J. Phys. Chem., 61, 562 (1957).

126. A. M. Michelson, The Chemistry of Nucleosides and Nucleotides (New York: Academic Press, 1963) pp. 445 ff.

127. A. M. Michelson and C. Monny, Proc. Nat. Acad. Sci., U. S., 56, 1528 (1966).

128. D. W. Miles, R. K. Robins and H. Eyring, Proc. Nat. Acad. Sci., U. S., 57, 1138 (1967).

129. W. G. Miller, D. A. Brant and P. J. Flory, J. Mol. Biol., 23, 67 (1967).

130. R. S. Morgan, Disc. Faraday Soc., 25, 193 (1958).

131. U. V. Morozov, Biofizika, 8, 331 (1963).

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

133. J. N. Murrell, The Theory of the Electronic Spectra of Organic Molecules (London: Metheun and Co., Ltd., 1963), p. 143.

134. H. A. Nash and D. F. Bradley, J. Chem. Phys., 45, 1380 (1966).

135. G. Nemethy and H. A. Scheraga, Biopolymers, 3, 155 (1965).

136. T. Ooi, R. A. Scott, R. F. Epand and H. A. Scheraga, J. Am. Chem. Soc., 88, 5680 (1966).

137. M. K. Pal. Histochemie, 5, 24 (1965).

138. M. K. Pal and M. Shubert, J. Phys. Chem., 67, 1821 (1963).

139. L. Pauling, The Nature of the Chemical Bond, 3rd ed. (Ithaca: Cornell University Press, 1960) p. 453.

140. L. Pauling and R. B. Corey, Fortschr. Chem. org. Naturstoffe, 11, 180 (1954).

141. M. F. Perutz, J. C. Kendrew and H. C. Watson, J. Mol. Biol., 13, 669 (1965).

142. D. C. Phillips, Proc. Nat. Acad. Sci., U. S., 57, 484 (1967).

143. K. S. Pitzer, J. Chem. Phys., 14, 239 (1946).

144. K. Pitzer, Adv. Chem. Phys., 2, 59 (1959).

145. K. S. Pitzer and W. E. Donath, J. Am. Chem. Soc., 81, 3213 (1959).

146. M. Pollak and R. Rein, J. Theor. Biol., 11, 490 (1966).

147. D. Poland, J. N. Vournakis and H. A. Scheraga, Biopolymers, 4, 223 (1966).

148. M. P. Printz and P. H. Von Hippel, Proc. Nat. Acad. Sci., U. S., 53, 362 (1965).

149. P. L. Privalov, K. A. Kafiani and D. R. Monaselidze, Biofizika, 10, 393 (1964).

150. A. Pullman, personal communication (1966).

151. B. Pullman, J. Chem. Phys., 43, S233 (1965).

152. B. Pullman, P. Claverie and J. Caillet, C. R. Acad. Sci., Paris, 263, 2006 (1966).

153. B. Pullman, P. Claverie and J. Caillet, Proc. Nat. Acad. Sci., U. S., 55, 904 (1966).

154. B. Pullman, P. Claverie and J. Caillet, Proc. Nat. Acad. Sci., U. S., 57, 1663 (1967).

155. E. Rabinowitch and L. F. Epstein, J. Am. Chem. Soc., 63, 69 (1947).

156. G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, J. Mol. Biol., 7, 95 (1963).

157. G. N. Ramachandran, C. M. Venkatachalam and S. Krimm, Biophys. J., 6, 849 (1966).

158. C. Ramakrishnan and G. N. Ramachandran, Biophys. J., 5, 909 (1965).

159. K. K. Reddi, Biochim. Biophys. Acta, 27, 1 (1958).

160. W. Rhodes, J. Am. Chem. Soc., 83, 3609 (1961).

161. A. Rich, D. R. Davies, F. H. C. Crick and J. D. Watson, J. Mol. Biol., 3, 71 (1961).

162. E. G. Richards, C. P. Flessel and J. R. Fresco, Biopolymers, 1, 431 (1963).

163. P. D. Ross and R. L. Scruggs, Biopolymers, 3, 491 (1965).

164. J. A. Rupley and V. Gates, Proc. Nat. Acad. Sci., U. S., 57, 496 (1967).

165. P. K. Sarkar and J. T. Yang, J. Biol. Chem. 240, 2088 (1965).

166. V. Sasisekharan, A. V. Lakshminarayanan and G. N. Ramachandran, personal communication (1966).

167. H. A. Scheraga, The Proteins, ed. H. Neurath (New York: Academic Press, 1963) v. I, pp. 477 ff.

168. P. R. Schimmel and P. J. Flory, Proc. Nat. Acad. Sci., U. S., 58, 52 (1967).

[illegible]. [illegible], K. T. Chen and E. O. P. Thilo,

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

170. R. A. Scott and H. A. Scheraga, J. Chem. Phys., 42, 2209 (1965).

171. R. A. Scott and H. A. Scheraga, J. Chem. Phys., 45, 2091 (1966).

172. E. Shefter, M. Barlow, R. Sparks and K.N. Trueblood, J. Am. Chem. Soc., 86, 1872 (1964).

173. E. Shefter and K. N. Trueblood, Acta Cryst., 18, 1067 (1965).

174. S. E. Sheppard, Proc. Roy. Soc. (London), A82, 256 (1909).

175. O. Sinanoglu and S. J. Abdulnur, Photochem. and Photobiol., 3, 333 (1964).

176. T. N. Solie, Thesis, University of Oregon, Eugene (1965).

177. M. Spencer, Acta Cryst., 12, 59 (1959).

178. M. Spencer, personal communication (1967).

178a E. Stellwagen and H. K. Schachman, Biochemistry, 1, 1056 (1962).

179. P. R. Stroesser and S. J. Gill, J. Phys. Chem. 71, 564 (1967).

180. M. Sundaralingam, J. Am. Chem. Soc., 87, 599 (1965).

181. M. Sundaralingam, Acta Cryst., 21, 495 (1966).

182. M. Sundaralingam and L.H. Jensen, J. Mol. Biol., 13, 914 (1965).

183. M. Sundaralingam and L. H. Jensen, J. Mol. Biol., 13, 960 (1965).

184. Y. K. Syrkin and M. E. Dyatkina, Structure of Molecules and the Chemical Bond, trans. M. A. Partridge and D. O. Jordan (New York: Dover, 1964) p. 199.

185. W. Szer, Biochem. Biophys. Res. Comm., 20, 182 (1965).

186. W. Szer and D. Shugar, J. Mol. Biol., 17, 174 (1966).

187. J. Thiery, J. Chim. Phys., in press.

188. R. Thomas, Biochim, Biophys. Acta, 14, 1231 (1954).

189. I. Tinoco, Jr., J. Am. Chem. Soc., 82, 4785 (1960); J. Am. Chem. Soc., 84, 5047 (1961).

190. I. Tinoco, Jr., Adv. Chem. Phys., 4, 113 (1962).

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

192. B. L. Tomlinson, personal communication (1967).

193. K. N. Trueblood, P. Horn and V. Luzzati, Acta Cryst., 14, 965 (1961).

194. P. O. P. Ts'o and S. I. Chan, J. Am. Chem. Soc., 86, 4176 (1964).

195. P. O. P. Ts'o, G. K. Helmkamp and C. Saunder, Proc. Nat. Acad. Sci., U. S., 48, 686 (1962).

196. P. O. P. Ts'o, I. S. Melvin, and A. C. Olsen, J. Am. Chem. Soc., 85, 1289 (1963).

197. P. O. P. Ts'o, S. A. Rapaport and F. J. Bollum, Biochemistry, 5, 4153 (1966).

198. P. O. P. Ts'o and M. P. Schweizer, personal communication (1967).

199. V. N. Tsvetkov, L. N. Andreeva and L. N. Kvitchenko, Vysokomol. Soed., 7, 2001 (1965).

200. V. G. Tumanian, N. G. Esipova and L. L. Kiselev, Dokl. Akad. Nauk, USSR, 168, 215 (1966).

201. T. L. V. Ulbricht, T. R. Emerson and R. J. Swan, Biochem. Biophys. Res. Comm., 19, 643 (1965).

202. G. Vanderkooi, S. J. Leach, G. Nemethy, R. A. Scott and H. A. Scheraga, Biochemistry, 5, 2991 (1966).

203. K. E. Van Holde, J. Brahms and A. M. Michelson, J. Mol. Biol., 12, 726 (1965).

204. K. E. Van Holde and G. P. Rosetti, Biochemistry, 6, 2189 (1967).

205. D. Voet, W. B. Gratzer, R. A. Cox and P. Doty, Biopolymers, 1, 13 (1963).

206. J. N. Vournakis, D. Poland and H. A. Scheraga, Biopolymers, 5, 403 (1967).

207. A. Wacker and E. Lodermann, Ang. Chemie (Int. Eng. Ed.) 4, 150 (1965).

208. M. M. Warshaw, Thesis, University of California, Berkeley (1965).

209. M. M. Warshaw, C. A. Bush and I. Tinoco, Jr., Biochem. Biophys. Res. Comm., 18, 633 (1965).

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

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

212. D. Watson, D. J. Sutor and P. Tolin, Acta Cryst., 19, 111 (1965).

213. J. D. Watson, Molecular Biology of the Gene (New York: W. A. Benjamin, Inc., 1965).

214. J. D. Watson and F. H. C. Crick, Nature, 171, 737 (1953).

215. J. D. Watson and F. H. C. Crick, Nature, 171, 964 (1953).

216. O. E. Weigang, J. Chem. Phys., 41, 1435 (1964).

217. D. E. Williams, J. Chem. Phys., 45, 3770 (1966).

218. R. W. Woody, M. E. Friedman and H. A. Scheraga, Biochemistry, 5, 2034 (1966).

219. J. T. Yang, T. Samejima and P. K. Sarkar, Biopolymers, 4, 623 (1966).

220. J. H. Young, Thesis, University of California, Berkeley, (1966).

221. V. Zanker, Z. Physik. Chem. (Leipzig), 199, 225 (1952).

222. V. Zanker, Z. Physik. Chem. (Leipzig), 200, 250 (1952).

# APPENDIX A

The following are some symbols and abbreviations used frequently in the text:

| | |
|---|---|
| A | Adenosine |
| U | Uridine |
| C | Cytidine |
| G | Guanosine |
| N | A general nucleoside |
| dN | Deoxynulceoside |
| Np | 3'(2') nucleotide |
| pN | 5' nucleotide |
| NpN | 3'-5' dinucleoside phosphate |
| poly N | Homopolynucleotide |
| poly (A+U) | 1:1 complex of poly A and poly U |
| DNA | Deoxyribonucleic acid |
| RNA | Ribonucleic acid |
| tRNA | Amino-acyl transfer RNA |
| ORD | Optical rotatory dispersion |
| CD | Circular dichroism |
| NMR | Nuclear magnetic resonance |
| $[\Phi]$ | Molar rotation per residue |
| $\kappa$ | Force constant of the torsional spring as defined in the oscillating dimer model |
| $T_m$ | Temperature of the midpoint of an ordered to disordered transition |
| $\phi$ | Torsional angle for rotation about a single covalent bond |
| $\phi_{CN}$ | Torsional angle for the glycosidic bond |
| H | Enthalpy |
| S | Entropy |
| F | Gibbs free energy |

# APPENDIX B

In this appendix we list some of the more useful programs and subroutines which we have used. They are all written in Chippewa Fortran for use with CDC 6000 series computers. These listed are:

| | |
|---|---|
| DELRE | Calculates the σ charge distribution by the method of Del Re. |
| LEMON | Calculates the torsional angles for rotations about bonds in known structures. |
| BRUTE | Calculates the energies of conformations generated by rotations of rigid groups about single covalent bonds. |
| Sburoutine DERIV | Calculates the first, second and third partial derivatives of the energy with respect to the torsional angles. |

```
      PROGRAM DELRE (INPUT,OUTPUT)
C     THIS PROGRAM COMPUTES THE CHARGE DISTRIBUTION OF A SIGMA SYSTEM
C     BY THE METHOD OF DEL RE.  (SEE 'ELECTRONIC ASPECTS OF BIOCHEMISTRY'
C     EDITED BY B. PULLMAN.  PAGES 221-235.)
C     THE PROGRAM IS WRITTEN IN CHIPPEWA FORTRAN.
C     THE MAXIMUM NUMBER OF ATOMS WHICH THIS PROGRAM CAN TREAT IS 40.
C     THE CHARGE DISTRIBUTION IS IN TERMS OF THE UNIT ELECTRON CHARGE.
C                              LIST OF VARIABLES USED
C     CHARGE(I)          CHARGE OF THE ITH ATOM
C     EPSILON(I,J)       DEL RE'S EPSILON
C     IQ                 IF IQ=0 ANOTHER COMPOUND IS TREATED.
C                        IF IQ IS NOT ZERO THEN THE PROGRAM EXITS.
C     N                  NUMBER OF ATOMS
C     NBRMAX(I)          NUMBER OF BONDED NEIGHBORS OF THE ITH ATOM
C     NHBR(I,J)          LABEL DESIGNATING THE ITH BONDED NEIGHBOR OF
C                        THE ITH ATOM.
C     Q1(I)              HOLLERITH FIELD DESCRIBING THE ITH ATOM.
C     Q2(I)              HOLLERITH FIELD DESCRIBING THE ITH ATOM.
C     QAB(I,J)           DEL RE'S CAPITAL Q
C     QSAB(I,J)          DEL RE'S LOWER CASE Q
C     V(I,J)             MINUS DEL RE'S GAMMA EXCEPT FOR J=N+1 IN WHICH
C                        CASE IT IS DEL RE'S DELTAO
C     X(I)               DEL RE'S DELTA
      0DIMENSION CHARGE(40), EPSILON(40,40), NBRMAX(40),
     1NHBR(40,4),Q1(40),Q2(40),QAB(40,40), QSAB(40,40),X(41),IL(41)
      COMMON/FLINC/V(40,41)
  115 CONTINUE
      PRINT 118
  118 FORMAT (1H1   )
      PRINT 117
 1170FORMAT (99H                    ATOM                                    CHA
     1RGE                                                          /// )
      READ 100, N, IQ
  100 FORMAT (2I2)
      DO 99 I=1,N
      DO 99 J=1,N
      V(I,J)=0.
   99 EPSILON(I,J)=10.**300
      K=N+1
      DO 103 I=1,N
C     READING IN ATOMIC DESCRIPTIONS AND PARAMETERS.
      READ 101, Q1(I), Q2(I), V(I,K), NBRMAX(I), (NHBR(I,J), J=1,4)
C     DESCRIPTION OF ATOM, DELTAO, NUMBER OF NEIGHBORS, LABELS OF
C     BONDED NEIGHBORS
  101 FORMAT (2A10, E10.3, 5I2)
      I1=NHBR(I,1)
      I2=NHBR(I,2)
      I3=NHBR(I,3)
      I4=NHBR(I,4)
      IF (NBRMAX(I)-3) 128, 125, 127
  125 CONTINUE
     0READ 116, V(I,I1), EPSILON(I,I1), V(I,I2), EPSILON(I,I2),
     1V(I,I3), EPSILON(I,I3)
  116 FORMAT (6E10.3)
      GO TO 103
```

```
  127 CONTINUE
      ORLAD 102, V(I,I1), EPSILON(I,I1), V(I,I2), EPSILON(I,I2),
     1V(I,I3), EPSILON(I,I3), V(I,I4), EPSILON(I,I4)
  102 FORMAT (8F10.3)
      GO TO 103
  128 CONTINUE
      IF (NBRMAX(I)-1) 119, 121, 123
  119 CONTINUE
      PRINT 120
  120 FORMAT ( 18H ERROR IN INPUT.      )
      GO TO 114
  121 CONTINUE
      READ 122, V(I,I1), EPSILON(I,I1)
  122 FORMAT (2F10.3)
      GO TO 103
  123 CONTINUE
      READ 124, V(I,I1), EPSILON(I,I1), V(I,I2), EPSILON(I,I2)
  124 FORMAT (4F10.3)
  103 V(I,I)=+1.0
C     SOLVING LINEAR SIMULTANEOUS EQUATIONS.
      M=1
      CALL FLIN(N,M,DET)
      K=N+1
      DO 1105 I=1,N
 1105 X(I)=V(I,K)
C     COMPUTING CHARGES.
      SCHARGE=0.0
      DO 112 I=1,N
      CHARGE(I)=0.0
      DO 110 J=1,N
      QAB(I,J)=(X(J)-X(I))/(2.*EPSILON(I,J))
      QSAB(I,J)=QAB(I,J)/(SQRTF(1.+(QAB(I,J)**2)))
  110 CHARGE(I)=CHARGE(I)+QSAB(I,J)
      PRINT 111, Q1(I), Q2(I), CHARGE(I)
  111 FORMAT ( 10X, 2A10, 20XF10.3)
  112 SCHARGE=SCHARGE+CHARGE(I)
      PRINT 130, SCHARGE
  130 FORMAT ( 20H THE TOTAL CHARGE =  E20.3)
      PRINT 113
  113 FORMAT (/  50H END OF CALCULATION FOR THIS COMPOUND.                )
      IF (IO) 114, 115, 114
  114 CONTINUE
      STOP
      END
      SUBROUTINE FLIN(JN,IM,DET)
      COMMON /FLINC/ V(40,41)
C                          MATRIX DIVISION
C     REPLACES B BY (A INVERSE)*B.  DESTROYS A.
C     A IS N BY N, B IS N BY M.  A AND B ARE STORED IN V(N,N+M).
      INTEGER P
      N=JN
      M=IM
C
      NM1=N-1
      NP1=N+1
```

```
      NPM=N+M
      DETERM=1.0
      P=1
  202 CONTINUE
      IMIN=P+1
C                                 FIND PIVOT
      PIVOT=0.
      DO 209 I=P,N
      SAVE=ABS(V(I,P))
      IF (SAVE .LE. PIVOT) GO TO 209
      PIVOT=SAVE
      IBIG=I
  209 CONTINUE
      IF (PIVOT .GT. 0.0  ) GO TO 210
      DETERM=0.0
      GO TO 290
C                                 ROW EXCHANGE AND ROW REDUCTION
  210 PIVOT=V(IBIG,P)
      DETERM=DETERM*PIVOT
      DO 219 J=P,NPM
      SAVE=V(IBIG,J)/PIVOT
      V(IBIG,J)=V(P,J)
      V(P,J)=SAVE
  219 CONTINUE
      IF(IBIG .NE. P) DETERM=-DETERM
      IF (P .GE. N) GO TO 250
      V(P,P)=1.0
C                                 BLOCK REDUCTION
      DO 239 I=IMIN,N
      DO 238 J=IMIN,NPM
      V(I,J)=V(I,J)-V(I,P)*V(P,J)
  238 CONTINUE
      V(I,P)=0.0
  239 CONTINUE
      P=P+1
  249 GO TO 202
C                                 BACK SUBSTITUTION
  250 CONTINUE
      IF (M .LE. 0) GO TO 290
      DO 259 K=NP1,NPM
      DO 253 P=1,NM1
      I=N-P
      DO 252 J=I,NM1
      V(I,K)=V(I,K)-V(J+1,K)*V(I,J+1)
  252 CONTINUE
  253 CONTINUE
  259 CONTINUE
C
  290 CONTINUE
      DET=DETERM
      RETURN
       END
```

DELRE input:

1st card — number of atoms, control number indicating if this is the last data set.

2nd card — description of atom, Del Re's $\delta$, number of nearest neighbors, identification numbers of the nearest neighbors.

3rd card — minus Del Re's $\gamma$, Del Re's $\epsilon$ (repeated for all nearest neighbors). Cards 2 and 3 are repeated for all atoms in the molecule.

```
15 1
ADENINE N1          +2.400E-01 2 2 6
-1.000E-01+1.000E+00-1.000E-01+1.000E+00
ADENINE C2          +7.000E-02 3 1 315
-1.000E-01+1.000E+00-1.000E-01+1.000E+00-3.000E-01+1.000E+00
ADENINE N3          +2.400E-01 2 2 4
-1.000E-01+1.000E+00-1.000E-01+1.000E+00
ADENINE C4          +7.000E-02 3 3 512
-1.000E-01+1.000E+00-1.000E-01+1.000E+00-1.000E-01+1.000E+00
ADENINE C5          +7.000E-02 3 4 6 7
-1.000E-01+1.000E+00-1.000E-01+1.000E+00-1.000E-01+1.000E+00
ADENINE C6          +7.000E-02 3 5 9 1
-1.000E-01+1.000E+00-1.000E-01+1.000E+00-1.000E-01+1.000E+00
ADENINE N7          +2.400E-01 2 5 8
-1.000E-01+1.000E+00-1.000E-01+1.000E+00
ADENINE C8          +7.000E-02 3 7 911
-1.000E-01+1.000E+00-1.000E-01+1.000E+00-3.000E-01+1.000E+00
ADENINE N9          +2.400E-01 3 6 810
-1.000E-01+1.000E+00-1.000E-01+1.000E+00-3.000E-01+4.500E-01
ADENINE H9          +0.000E+00 1 9
-4.000E-01+4.500E-01
ADENINE H8          +0.000E+00 1 8
-4.000E-01+1.000E+00
ADENINE AMINO N     +2.400E-01 3 41413
-1.000E-01+1.000E+00-3.000E-01+4.500E-01-3.000E-01+4.500E-01
ADENINE AMINO H     +0.000E+00 112
-4.000E-01+4.500E-01
ADENINE AMINO H     +0.000E+00 112
-4.000E-01+4.500E-01
ADENINE H2          +0.000E+00 1 2
-4.000E-01+1.000E+00
(
```

```
      PROGRAM LEMON (INPUT,OUTPUT)
C     THIS PROGRAM WAS WRITTEN IN CHIPPEWA FORTRAN.
C     GIVEN EITHER CYLINDRICAL OR CARTESIAN COORDINATES IT COMPUTES
C     THE ORIENTATION ANGLE ABOUT ANY BOND.  THE ANGLE IS MEASURED
C     FROM THE CONFORMATION WHERE THE TWO ORIENTING ATOMS ARE
C     CIS-COPLANAR.
C     ICYLIND=0 THEN THE COORDINATES ARE CYLINDRICAL, IF NOT, CARTESIAN.
C     MAXATOM IS THE NUMBER OF ATOMS IN THE INPUT.    MAXIMUM - 100 ATOMS
C     MAXBOND IS THE NUMBER OF BONDS IN THE INPUT.    NO MAXIMUM
C     COORD(ATOMNUMBER,I) ARE THE COORDINATES OF THE ATOMS EITHER X,Y,Z
C     OR R,THETA,(IN DEGREES), Z.
C     ATOMLBL IS A HOLLERITH VARIABLE DESCRIBING THE ATOM (URACIL C6)
C     STRUC   IS A HOLLERITH VARIABLE DESCRIBING THE MOLECULE (B DNA)
C     JATOMS ARE THE ATOMS DEFINING THE BOND- 1 AND 2 AT EITHER END OF
C     THE BOND. 3 AND 4 DEFINING ITS ORIENTATION.
C     ATOM 1 IS BONDED TO ATOM 4, ATOM 2 TO ATOM 3
      DIMENSION STRUC(7), ATOMLBL(100,2),    COORD(100,9)
   13 READ 1, ICYLIND, MAXATOM, MAXBOND, (STRUC(I),  I=1,7)
    1 FORMAT (3I3, 7A10)
      DO 2 L=1,MAXATOM
    2 READ 3, ATOMLBL(L,1), ATOMLBL(L,2), (COORD(L,I), I=1,3)
    3 FORMAT (2A10, 3E11.4)
      IF (ICYLIND) 5,4,5
C     CONVERTING TO CARTESIAN COORDINATES IF NECESSARY -
    4 DO 6 L=1,MAXATOM
      THETA=COORD(L,2)*3.141592654/180.
      COORD(L,2)=COORD(L,1)*SIN(THETA)
    6 COORD(L,1)=COORD(L,1)*COS(THETA)
    5 PRINT 7, (STRUC(I), I=1,7)
    7 FORMAT (1H1 33H THE STRUCTURE BEING ANALYZED IS    7A10    ///)
      DO 8 JBOND=1,MAXBOND
      READ 9, JATOM1, JATOM2, JATOM3, JATOM4
    9 FORMAT (4I3)
C     ALLIGNING BOND ALONG THE Y AXIS
C     FIRST TRANSLATING TO THE ORIGIN
      DO 10 I=1,3
      J=I+3
      COORD(JATOM2,J)=COORD(JATOM2,I)-COORD(JATOM1,I)
      COORD(JATOM3,J)=COORD(JATOM3,I)-COORD(JATOM1,I)
   10 COORD(JATOM4,J)=COORD(JATOM4,I)-COORD(JATOM1,I)
C     ROTATING
      A=SQRT(COORD(JATOM2,5)**2+COORD(JATOM2,4)**2)
      B=SQRT(A**2+COORD(JATOM2,6)**2)
      CZ=COORD(JATOM2,5)/A
      SZ=COORD(JATOM2,4)/A
      SX=COORD(JATOM2,6)/B
      CX=A/B
      COORD(JATOM3,7)=COORD(JATOM3,4)*CZ-SZ*COORD(JATOM3,5)
     0COORD(JATOM3,9)=-SZ*SX*COORD(JATOM3,4)-CZ*SX*COORD(JATOM3,5)+
     1CX*COORD(JATOM3,6)
      C=SQRT(COORD(JATOM3,7)**2+COORD(JATOM3,9)**2)
      CY=COORD(JATOM3,7)/C
      SY=COORD(JATOM3,9)/C
     0XCOORD=COORD(JATOM4,4)*(CY*CZ-SX*SY*SZ)-
     1COORD(JATOM4,5)*(CY*SZ+SX*SY*CZ)+COORD(JATOM4,6)*(CX*SY)
```

```
      ZCOORD=-COORD(JATOM4,4)*(CZ*SY+SX*CY*SZ)+
     1COORD(JATOM4,5)*(SY*SZ-SX*CY*CZ)+COORD(JATOM4,6)*CX*CY
      ANGLE=ATAN2(ZCOORD,XCOORD)*180./3.141592654
      WRITE 11, (ATOMLBL(JATOM1,I), I=1,2), (ATOMLBL(JATOM2,I), I=1,2),
     1(ATOMLBL(JATOM3,I), I=1,2), (ATOMLBL(JATOM4,I), I=1,2)
   11 FORMAT (/27H THE ORIENTATION ABOUT THE 2A10, 3H - 2A10, 15H AS DEF
     1INED BY 2A10, 5H AND 2A10                                        )
      WRITE 12, ANGLE
   12 FORMAT ( 4H IS F20.7, 35H DEGREES COUNTER-CLOCKWISE.              )
      GO TO 10
      STOP
      END
```

LEMON input:

| | |
|---|---|
| 1st card | ICYLIND=0 if the coordinates are Cartesian, $\neq$ 0 if the coordinates are cylindrical, number of atoms, number of bonds, name of structure. |
| 2nd card | Atomic description, atomic coordinates (in Å and/or degrees) This format is repeated for all atoms. |
| last cards | 1st atom on bond, 2nd atom on bond, atom connected to the 2nd (used to define standard orientation), atom connected to the 1st (used to define standard orientation) This format is repeated for all bonds to be considered. |

```
1     4   1 2' ENDO RIBOSE FROM CYTIDYLIC ACID
2' ENDO RIBOSE C2'  -6.2310E-01-5.8760E-01-1.2480E+00
2' ENDO RIBOSE O2'  -2.0250E+00-5.6670E-01-1.2640E+00
2' ENDO RIBOSE OH2' -2.6060E+00-1.0950E+00-5.9350E-01
2' ENDO RIBOSE C1'  +0.0000E+00+0.0000E+00+0.0000E+00
  1   2   3   4
'
```

```
      PROGRAM BRUTE (INPUT,OUTPUT,TAPE1)
C     PROGRAM BRUTE COMPUTES THE ENERGIES OF VARIOUS CONFIGURATIONS
C     OF DINUCLEOSIDE PHOSPHATES.  TERMS INCLUDED ARE CHARGE-CHARGE,
C     CHARGE-POLARIZABILITY AND A 6-12 POTENTIAL.  AN EFFORT IS MADE TO
C     SPEED THE CALCULATION BY THROWING OUT CONFIGURATIONS IN TWO
C     GROUNDS.  1-THEY WERE ELIMINATED IN EARLIER WORK ON PARTS OF THE
C     DINUCLEOSIDE PHOSPHATE.  THIS CAN BE DONEBY EITHER NOT INCLUDING
C     THE CONFIGURATION TO BEGIN WITH OR SORTING IT OUT IN SUBROUTINE
C     SORT, 2- THROWING OUT CONFIGURATIONS IN WHICH THE INTERATOMIC
C     DISTANCES ARE LESS THAN 62 PER CENT OF THE SUM OF THE VAN DER
C     WAALS RADII.
C     OUTPUT OF ENERGIES IS ON MAGNETIC TAPE WITH APPROXIMATELY 25000
C     RECORDS PER REEL.
C     SUBROUTINES USED ARE --
C     INTAKE -            READS IN DATA
C     POSTN -             ROTATES ABOUT THE BOND OF INTEREST
C     SURFACE -           COMPUTES THE ADDITIONAL ENERGY CAUSED BY
C                   INTERACTIONS BETWEEN TWO SIDES OF A BOND.
C                   IT SHOULD BE NOTED THAT IN ORDER FOR THIS SCHEME TO
C                   WORK THE BONDS MUST BE IN LINEAR ORDER AS THEY APPEAR
C                   IN THE MOLECULE.
C     SORT -              SORTS OUT THE UNWANTED CONFIGURATIONS.
C     EXHAUST             READS OUT THE FINAL DATA.
C     SETUP         READS IN PARAMETERS AND INITUALIZES THE BOTTOM
C                   ARRAY.
C     GRAPHIC       PICKS THE MINIMUM ENERGY FOR EACH POINT ON THE
C                   TWO DIMENSIONAL SURFACE.
C     VIEWER        PRINTS OUT THE TWO DIMENSIONAL SURFACE WHERE THE
C                   FUNCTION IS THE LOWEST ENERGY OF A CONFORMATION FOR
C                        THE TWO SPECIFIED TORSIONAL ANGLES AND ANY VALUE
C                        CALCULATED FOR THE OTHER ANGLES.
C     VIEWER        PRINTS OUT THE SURFACE.
C     VARIABLES USED IN THIS CALCULATION ARE -
C     THOSE NEXT TO AN * ARE INPUT DATA
C     A                   EITHER THE R6 COEFFICIENT OR THE DISTANCE FROM
C                   THE Z AXIS OF AN ATOM ON THE END OF A BOND.
C     A11,A12,...         MATRIX ELEMENTS FOR THE ROTATION MATRIX
C     AGROUP(I,J,K)       AT ONE TIME THE X,Y,ANDZ COORDINATES OF THE JTH
C                   ATOM OF THE ITH GROUP
C    *ANGLE(I)            THE ANGLE WHICH POSTN ACTUALLY ROTATES ABOUT
C                   THE ITH BOND.   IT MAY EITHER BE THE SUM OF ONE OR MORE
C                   THETA(I,J) OR IT MAY BE THE ORIGINAL ORIENTATION OF
C                   THE BOND.
C     B                   EITHER THE R12 COEFFICIENT OR THE DISTANCE FROM
C                   THE X AXIS (SEE A)
C     BOTTOM(I,J,K)        IS THE MINIMUM  VALUE OF THE ENERGY  FOR A
C                   GIVEN SET OF ANGLES WHERE WE CONSIDER THE JTH ANGLE
C                   FOR THE FIRST BOND AND THE KTH FOR THE SECOND ON THE
C                   ITH SURFACE.  THERE CAN BE UP TO 2  SURFACES.
C     CX,CZ               COSINES OF THE ANGLES OF ROTATION ABOUT THE X
C                   AND Z AXES
C     DMIN                65 PERCENT OF THE VAN DER WAALS DISTANCE
C     DTHETA(I)           THE CUMULATIVE ANGLE OF ROTATION (IN DEGREES)
C                   ABOUT THE ITH BOND.
C     E1,E2,E3            PAIRWISE INTERATOMIC CHARGE-CHARGE, CHARGE-
```

```
C                    POLARIZABILITY AND 6-12 INTERACTIONS.
C     ELECN1,ELECN2        'EFFECTIVE NUMBER OF ELECTRONS' OF AN ATOM
C     E                    THE ENERGY OF INTERACTION WHICH CAN BE CHANGED
C                    BY ROTATING AROUND A SINGLE BOND.
C     E1,E2,E3             RUNNING TOTALS OF E1,E2,E3
C    *GROUP(I,J,K)         THE COORDINATES,CHARGE,POLARIZABILITY AND ATOM
C                    INDEX OF THE JTH ATOM OF THE ITH GROUP.
C     I                    THE INDEX FOR THE ANGLE OF ROTATION
C     I1,I2,...            SAME AS I BUT FOR THE INDICATED ANGLE
C     ICOUNT               A COUNTER INDICATING THE NUMBER OF CONFORMATIONS
C                    READ OUT.
C     IFIRST,ILAST         THE FIRST AND LAST GROUPS MOVED WHEN A BOND IS
C                    ROTATED.
C     II                   GROUP NUMBER AT THE END OF THE BOND NOT ATTACHED
C                    TO ATOMS WHICH ARE TO BE ROTATED.
C    *INDEX(IPLOT,I)        ARE THE INDECES OF THE IPLOT TH SURFACE
C                    I=1 IS THE FIRST BOND AND I=2 IS THE SECOND.
C     IROTN(I,J)           REFERS TO THE ITH BOND TO BE ROTATED.
C                          J=1, GROUP OF THE ATOM AT THE OTHER END.
C                          J=2, NUMBER OF THE ATOM AT THE OTHER END.
C                          IROTN(I,3) BEING THE FIRST AND IROTN(I,4) BEING
C                          THE LAST.
C     ISTOP                SAME AS II
C     IZ                   LABEL FOR THE BOND NUMBER
C     J                    VARIABLE USED IN SUMMING OVER X,Y,Z POSITIONS
C     JATOM1,JATOM2        NUMBER OF ATOM1 OR ATOM2
C     JGRP1,JGRP2M2        NUMBER OF THE FIRST OR SECOND GROUP
C     JJ                   ATOM NO. AT THE END OF THE BOND NOT TO BE MOVED.
C     JPRIME               A DODGE SO THAT I CAN SUM OVER GROUPS IN THE
C                    REVERSE ORDER.
C     JUMP                 VARIALBE WHICH IF 0 INDICATES THAT THE
C                    CONFIGURATION UNDER CONSIDERATION IS TO BE ABORTED.
C     K                    A VARIABLE USED TO DO OVER A SET OF ATOMS
C     KK                   NUMBER OF ATOMS IN THE GROUP UNDER CONSIDERATION
C     L                    LABEL FOR GROUP NUMBER
C     MAXAT(I)             NUMBER OF ATOMS IN THE ITH GROUP  MAXIMUM 20
C     MAXAT1,MAXAT2        NUMBER OF ATOMS IN GROUPS ONE OR TWO
C     MAX1,MAX2,...        NUMBER OF ANGLES TO BE TRIED COR THE ITH BOND
C                    MAXIMUM 36
C     MAXGRP               NUMBER OF GROUPS OF ATOMS IN THE CALCULATION
C                    MAXIMUM 9
C    *MAXPLOT              IS THE NUMBER OF TWO DIMENSIONAL SURFACES OF
C                    BOTTOM TO BE PLOTTED OUT.      MAXIMUM 20
C     R                    INTERATOMIC DISTANCE
C     ROTSIN,ROTCOS        THE SIN AND COSINE OF THE ANGLE OF ROTATION
C                    ABOUT THE BOND UNDER CONSIDERATION)
C                    ABOUT THE BOND UNDER CONSIDERATION.
C     SX,SZ                SINES OF THE ROTATIONS ABOUT THE X AND Z AXES
C                    USED IN LINING A BOND UP ALONG THE Y AXIS
C    *THETA(I,J)           THE JTH ANGLE OF ROTATION ABOUT THE ITH BOND
C                    IN DEGREES.  NOTE THAT THESE ARE INCREMENTAL ANGLES.
C     NOTE THAT ONE CRITERIA WHICH MUST BE MET IN THIS PROGRAM IS THAT
C     ANGLES MUST COMPLETE A CYCLE WHICH RETURNS THEMSELVES TO THEIR
C     ORIGINAL POSITIONS.      (0 DEGREES)
C     TOTALEN(I)           THE ENERGY TERMS WHICH DON'T CHANGE UPON
```

```
C                    A CHANGE IN ANGLE I
C     TOTAL1,TOTAL2,TOTAL3   PARTS OF TOTALEN =  THE CHARGE-CHARGE,
C                    CHARGE-POLARIZABILITY AND 6-12 POTENTIALS
C     VWDIAM                 VAN DER WAALS DISTANCE
C     VWRAD1,VWRAD2          VAN DER WAALS RADII
C     X                      THETA IN RADIANS
      COMMON/BLOCKA/MAX1,MAX2,MAX3,MAX4,MAX5,MAX6,MAX7,MAX8
      COMMON/BLOCKB/THETA(8,36)
      COMMON/BLOCKD/ICOUNT,I1,I2,I3,I4,I5,I6,I7,I8
      COMMON/BLOCKE/DTHETA(8)
      COMMON/BLOCKG/JUMP
      COMMON/BLOCKH/I
      COMMON/BLOCKJ/IZ
      COMMON/BLOCKK/ANGLE(8)
 9999 CONTINUE
      ICOUNT=0
      CALL INTAKE
      DO 1 I1=1,MAX1
      IZ=1
      I=I1
      ANGLE(1)=THETA(1,I)
      DTHETA(1)=DTHETA(1)+ANGLE(1)
      CALL POSTN
      CALL SURFACE
      IF (JUMP.EQ.0) GO TO 1
      DO 2 I2=1,MAX2
      IZ=2
      I=I2
      DTHETA( 2)=DTHETA( 2)+THETA( 2,I)
      ANGLE(2)=THETA(2,I)
      CALL POSTN
      CALL SURFACE
      IF (JUMP.EQ.0) GO TO 2
      DO 3 I3=1,MAX3
      IZ=3
      I=I3
      DTHETA( 3)=DTHETA( 3)+THETA( 3,I)
      ANGLE(3)=THETA(3,I)
      CALL POSTN
      CALL SURFACE
      IF (JUMP.EQ.0) GO TO 3
      DO 4 I4=1,MAX4
      IZ=4
      I=I4
      DTHETA( 4)=DTHETA( 4)+THETA( 4,I)
      CALL SORT
      IF (JUMP.EQ.0) GO TO 4
      CALL POSTN
      CALL SURFACE
      IF (JUMP.EQ.0) GO TO 4
      DO 5 I5=1,MAX5
      IZ=5
      I=I5
      DTHETA( 5)=DTHETA( 5)+THETA( 5,I)
      CALL SORT
```

```
      IF (JUMP.EQ.0) GO TO 5
      CALL POSTN
      CALL SURFACE
      IF (JUMP.EQ.0) GO TO 5
      DO 6 I6=1,MAX6
      IZ=6
      I=I6
      DTHETA( 6)=DTHETA( 6)+THETA( 6,I)
      CALL SORT
      IF (JUMP.EQ.0) GO TO 6
      CALL POSTN
      CALL SURFACE
      IF (JUMP.EQ.0) GO TO 6
      DO 7 I7=1,MAX7
      IZ=7
      I=I7
      DTHETA( 7)=DTHETA( 7)+THETA( 7,I)
      CALL SORT
      IF (JUMP.EQ.0) GO TO 7
      CALL POSTN
      CALL SURFACE
      IF (JUMP.EQ.0) GO TO 7
      DO 8 I8=1,MAX8
      IZ=8
      I=I8
      DTHETA( 8)=DTHETA( 8)+THETA( 8,I)
      ANGLE(8)=THETA(8,I)
      CALL POSTN
      CALL SURFACE
      IF (JUMP.EQ.0) GO TO 8
      CALL EXHAUST
    8 CONTINUE
    7 CONTINUE
    6 CONTINUE
    5 CONTINUE
    4 CONTINUE
    3 CONTINUE
    2 CONTINUE
    1 CONTINUE
      PRINT 99,ICOUNT
      WRITE (1,99) ICOUNT
   99 FORMAT (// 31H THE NUMBER OF DATA ON TAPE IS I10)
      CALL VIEWER
      END FILE 1
      GO TO 9999
      STOP
      END
      SUBROUTINE INTAKE
C     READS IN DATA
      DIMENSION DPRIME(37)
      COMMON/BLOCKA/MAX1,MAX2,MAX3,MAX4,MAX5,MAX6,MAX7,MAX8
      COMMON/BLOCKB/THETA(8,36)
      COMMON/BLOCKE/DTHETA(8)
      COMMON/BLOCKF/GROUP(9,20,6),MAXAT(9),IROTN(8,4)
      COMMON/BLOCKG/JUMP
```

```
      COMMON/BLOCKJ/IZ
      COMMON/BLOCKK/ANGLE(8)
      COMMON/BLOCKL/MAXGRP
C     VARIABLES TO BE READ IN ARE AS FOLLOWS -
C     Q1 AND Q2 - HOLLENITH VARIABLES DESCRIBING ATOMS (E.G. URACIL C6.
C     GROUP(I,J,1), GROUP(I,J,2) AND GROUP(I,J,3. - THE X, Y, AND Z
C        COORDINATES OF THE JTH ATOM IN THE ITH GROUP (IN ANGSTROMS)
C     GROUP(I,J,4) - THE CHARGE OF THE JTH ATOM OF THE ITH GROUP
C        (IN ELECTRON CHARGES)
C     GROUP(I,J,5) - THE POLARIZABILITY OF THE JTH ATOM OF THE ITH GROUP
C        (IN CUBIC ANGSTROMS)
C     GROUP(I,J,6) - THE CODE NUMBER OF THE JTH ATOM IN THE ITH GROUP
C        WHICH IDENTIFIES THE TYPE OF ATOM (E.G. CARBON IS 2.0)
  100 FORMAT (1H1)
  101 FORMAT (4I2,2A10,F8.2)
  102 FORMAT (6H BOND I1, 18H IS IDENTIFIED AS 2A10)
  103 FORMAT (10F8.1)
  104 FORMAT (28H+ANGLES TO BE TRIED FOR BOND I2,5H ARE    )
  105 FORMAT (35X,9F10.1)
  106 FORMAT (55H ROTATION DOES NOT RETURN TO ORIGINAL POSITION FOR BOND
     1      I2)
   10 FORMAT (    98H1                                        -  -  -  INITIAL COO
     1RDINATES  -  -  -                                        )
   20 FORMAT ( 98H GROUP ATOM   IDENTIFICATION          X              Y
     1          Z            CHARGE    POLARIZABILITY                        )
    3 FORMAT (9I2)
    4 FORMAT (1H )
    5 FORMAT (26I3)
    7 FORMAT (2A10,5E11.3,F4.1)
   60 FORMAT ( 1X, I3, 3X, I3, 1X, 2A10, 1X, E11.3, 2X, E11.3, 2X, E11.3
     1, 2X, E11.3, 2X, E11.3)
      PRINT 100
      PRINT 1
      WRITE (1,1)
      PRINT 2
      WRITE (1,2)
      READ 3, MAXGRP,MAX1,MAX2,MAX3,MAX4,MAX5,MAX6,MAX7,MAX8
      READ 5,(MAXAT(JGRP),JGRP=1,MAXGRP)
      DO 8 JGRP=1,MAXGRP
      MAX=MAXAT(JGRP)
      PRINT 4
      WRITE (1,4)
      DO 8 JATOM =1,MAX
      READ 7,Q1,Q2,(GROUP(JGRP,JATOM,IQ),IQ=1,6)
      PRINT 6,      JGRP,JATOM,Q1,Q2,(GROUP(JGRP,JATOM,IQ),IQ=1,5)
    8 WRITE (1,6)   JGRP,JATOM,Q1,Q2,(GROUP(JGRP,JATOM,IQ),IQ=1,5)
      PRINT 4
      WRITE (1,4)
      I=1
      DO 9 IZ=1,8
      READ 101, (IROTN(IZ,IQ),IQ=1,4),Q1,Q2,ANGLE(IZ)
      DTHETA(IZ)=ANGLE(IZ)
      DPRIME(I)=ANGLE(IZ)
      PRINT 102, IZ,Q1,Q2
      WRITE (1,102) IZ,Q1,Q2
```

```
C       ALLIGNING MOLECULE FOR START OF CALCULATION.
      CALL POSTN
      GO TO (10,11,12,13,14,15,16,17),IZ
   10 READ 103, (THETA(IZ,I),I=1,MAX1)
      DO 201 I=1,MAX1
      IPLUS1=I+1
  201 DPRIME(IPLUS1)=DPRIME(I)+THETA(1,I)
      WRITE(1,104) IZ
      PRINT 104,IZ
      M=MAX1+1
      PRINT 105,(DPRIME(I),I=2,M)
      WRITE (1,105) (DPRIME(I),I=2,M)
      IF (DPRIME(M)-DPRIME(1)) 18,9,18
   18 PRINT 106,IZ
      WRITE (1,106) IZ
      STOP
   11 READ 103, (THETA(IZ,I),I=1,MAX2)
      DO 202 I=1,MAX2
      IPLUS1=I+1
  202 DPRIME(IPLUS1)=DPRIME(I)+THETA(2,I)
      WRITE(1,104) IZ
      PRINT 104,IZ
      M=MAX2+1
      PRINT 105,(DPRIME(I),I=2,M)
      WRITE (1,105) (DPRIME(I),I=2,M)
      IF (DPRIME(M)-DPRIME(1)) 18,9,18
   12 READ 103, (THETA(IZ,I),I=1,MAX3)
      DO 203 I=1,MAX3
      IPLUS1=I+1
  203 DPRIME(IPLUS1)=DPRIME(I)+THETA(3,I)
      WRITE(1,104) IZ
      PRINT 104,IZ
      M=MAX3+1
      PRINT 105,(DPRIME(I),I=2,M)
      WRITE (1,105) (DPRIME(I),I=2,M)
      IF (DPRIME(M)-DPRIME(1)) 18,9,18
   13 READ 103, (THETA(IZ,I),I=1,MAX4)
      DO 204 I=1,MAX4
      IPLUS1=I+1
  204 DPRIME(IPLUS1)=DPRIME(I)+THETA(4,I)
      WRITE(1,104) IZ
      PRINT 104,IZ
      M=MAX4+1
      PRINT 105,(DPRIME(I),I=2,M)
      WRITE (1,105) (DPRIME(I),I=2,M)
      IF (DPRIME(M)-DPRIME(1)) 18,9,18
   14 READ 103, (THETA(IZ,I),I=1,MAX5)
      DO 205 I=1,MAX5
      IPLUS1=I+1
  205 DPRIME(IPLUS1)=DPRIME(I)+THETA(5,I)
      WRITE(1,104) IZ
      PRINT 104,IZ
      M=MAX5+1
      PRINT 105,(DPRIME(I),I=2,M)
      WRITE (1,105) (DPRIME(I),I=2,M)
```

```
      IF (DPRIME(M)-DPRIME(1)) 18,9,18
   15 READ 103, (THETA(IZ,I),I=1,MAX6)
      DO 206 I=1,MAX6
      IPLUS1=I+1
  206 DPRIME(IPLUS1)=DPRIME(I)+THETA(6,I)
      PRINT 104,IZ
      WRITE(1,104) IZ
      M=MAX6+1
      PRINT 105,(DPRIME(I),I=2,M)
      WRITE (1,105) (DPRIME(I),I=2,M)
      IF (DPRIME(M)-DPRIME(1)) 18,9,18
   16 READ 103, (THETA(IZ,I),I=1,MAX7)
      DO 207 I=1,MAX7
      IPLUS1=I+1
  207 DPRIME(IPLUS1)=DPRIME(I)+THETA(7,I)
      PRINT 104,IZ
      WRITE(1,104) IZ
      M=MAX7+1
      PRINT 105,(DPRIME(I),I=2,M)
      WRITE (1,105) (DPRIME(I),I=2,M)
      IF (DPRIME(M)-DPRIME(1)) 18,9,18
   17 READ 103, (THETA(IZ,I),I=1,MAX8)
      DO 208 I=1,MAX8
      IPLUS1=I+1
  208 DPRIME(IPLUS1)=DPRIME(I)+THETA(8,I)
      WRITE(1,104) IZ
      PRINT 104,IZ
      M=MAX8+1
      PRINT 105,(DPRIME(I),I=2,M)
      WRITE (1,105) (DPRIME(I),I=2,M)
      IF (DPRIME(M)-DPRIME(1)) 18,9,18
    9 CONTINUE
      JUMP=1
      CALL SETUP
      RETURN
      END
      SUBROUTINE SORT
      COMMON/BLOCKB/THETA(8,36)
      COMMON/BLOCKE/DTHETA(8)
      COMMON/BLOCKG/JUMP
      COMMON/BLOCKH/I
      COMMON/BLOCKJ/IZ
      COMMON/BLOCKK/ANGLE(8)
      JUMP=1
      ANGLE(IZ)=ANGLE(IZ)+THETA(IZ,I)
      GO TO (100,2,3,4,5,6,7,8) IZ
    2 CONTINUE
      GO TO 100
    3 CONTINUE
      GO TO 100
    4 DTHETA3=DTHETA(3)
      DTHETA4=DTHETA(4)
      IF (DTHETA3.GT.(120.).AND.DTHETA3.LT.(300.)) GO TO 100
      IF (DTHETA3-180.)401,402,403
  403 IF(DTHETA3+DTHETA4-560.) 430,99,431
```

```
431 IF (DTHETA4.LE.(300.)) GO TO 99
    IF (DTHETA3+DTHETA4-660.) 100,99,99
430 IF (DTHETA4.LE.(60.)) GO TO 435
    IF (( DTHETA3+.5*DTHETA4).LT.(380.)) GO TO 100
    IF ((DTHETA3-DTHETA4).LT.(140.)) GO TO 100
    IF (DTHETA3+DTHETA4-500.) 99,99,100
435 IF ((DTHETA4+DTHETA3).LE.(380.).AND.(DTHETA3+DTHETA4).GE.(320.))
   160 TO 99
    GO TO 100
402 IF (DTHETA4.LE.(340.).AND.DTHETA4.GE.(260.)) GO TO 99
    IF (DTHETA4.LE.(220.).AND.DTHETA4.GE.(150.)) GO TO 99
    IF (DTHETA4.LE.( 90.).AND.DTHETA4.GE.( 20.)) GO TO 99
    GO TO 100
401 IF (DTHETA3-30.) 4000,4000,4001
4001 IF (DTHETA3-70.) 100,100,4400
4400 IF (DTHETA3-90.) 4401,4401,99
4401 IF (DTHETA4.GE.(20.).AND.DTHETA4.LE.(100.)) GO TO 99
    IF (DTHETA4.GE.(140.).AND.DTHETA4.LE.(210.)) GO TO 99
    IF (DTHETA4.GE.(270.).AND.DTHETA4.LE.(350.)) GO TO 99
    GO TO 100
4000 IF (DTHETA4-200.) 4003,99,4004
4004 IF (DTHETA4.GE.(300.)) GO TO 99
    IF ((2.*DTHETA4/3.+DTHETA3).LE.(180.)) GO TO 99
    GO TO 100
4003 IF (3.*DTHETA3+DTHETA4-150.) 4010,4010,100
4010 IF (DTHETA3-20.) 4011,4011,100
4011 IF (DTHETA4-60.) 4012,99,99
4012 IF (DTHETA3+DTHETA4-20.) 99,99,100
    GO TO 100
  5 DTHETA5=DTHETA(5)
    IF (DTHETA5-40.) 500,100,501
501 IF (DTHETA5-290.) 100,100,510
510 IF (DTHETA4-70.) 520,100,521
520 IF (DTHETA4-DTHETA5+270.) 99,99,100
521 IF (DTHETA4+DTHETA5-660.) 100,99,99
500 IF ((DTHETA4+DTHETA5).LE.(60.)) GO TO 99
    IF  (DTHETA4-DTHETA5-300.) 100,99,99
    GO TO 100
  6 DTHETA6=DTHETA(6)
    IF (DTHETA6-40.) 600,100,601
601 IF (DTHETA6-320.) 100,100,602
602 IF (DTHETA5.LE.(30.)) GO TO 99
    IF (DTHETA5.GE.(120.).AND.DTHETA5.LE.(150.)) GO TO 99
    GO TO 100
600 IF (DTHETA5-150.) 604,99,100
604 IF (DTHETA5-DTHETA6-30.) 605,99,99
605 IF (DTHETA5+DTHETA6-30.) 99,99,100
    GO TO 100
  7 DTHETA7=DTHETA(7)
    IF (DTHETA6-100.) 700,700,750
750 IF (DTHETA6.LT.(260.)) GO TO 100
    IF (DTHETA7-90.) 775,780,780
775 IF (DTHETA6-2.*DTHETA7-140.) 100,99,99
780 IF (DTHETA6+DTHETA7-620.) 781,99,99
781 IF (2.*DTHETA7-DTHETA6-240.) 782,782,100
```

```
  82 IF ((2.*DTHETA7+DTHETA6).GE.(800.)) GO TO 99
     IF (DTHETA6.LT.(340.)) GO TO 100
     IF (DTHETA7.LT.(200.)) GO TO 100
     GO TO 99
 700 IF (DTHETA7-DTHETA6-200.) 701,99,99
 701 IF (DTHETA6+DTHETA7-80.) 99,99,100
     GO TO 100
   8 CONTINUE
     GO TO 100
  99 JUMP=0
 100 RETURN
     END
     SUBROUTINE POSTN
     DIMENSION AGROUP(9,20,3)
     COMMON/BLOCKF/GROUP(9,20,6),MAXAT(9),IROTN(8,4)
     COMMON/BLOCKH/I
     COMMON/BLOCKJ/IZ
     COMMON/BLOCKK/ANGLE(8)
     X=ANGLE(IZ)*3.141592654/180.
     IFIRST=IROTN(IZ,3)
     ILAST=IROTN(IZ,4)
     II=IROTN(IZ,1)
     JJ=IROTN(IZ,2)
     DO 200 L=IFIRST,ILAST
     KK=MAXAT(L)
     DO 200 K=1,KK
     DO 200 J=1,3
C    TRANSLATING TO THE ORIGIN
 200 AGROUP(L,K,J)=GROUP(L,K,J)-GROUP(II,JJ,J)
C    ROTATING
     A=SQRT(AGROUP(IFIRST,1,1)**2+AGROUP(IFIRST,1,2)**2)
     CZ=AGROUP(IFIRST,1,2)/A
     SZ=AGROUP(IFIRST,1,1)/A
     B=SQRT(A**2+AGROUP(IFIRST,1,3)**2)
     CX=A/B
     SX=AGROUP(IFIRST,1,3)/B
     ROTCOS=COS(X)
     ROTSIN=SIN(X)
     A11=ROTCOS*(CZ**2)+ROTCOS*((SZ*SX)**2)+(CX*SZ)**2
     A12=-ROTSIN*SX-ROTCOS*SZ*CZ*(CX**2)+          CZ*SZ*(CX**2)
     A13=CZ*CX*ROTSIN+CX*SX*SZ-SZ*CX*SX*ROTCOS
     A21=-ROTCOS*SZ*CZ*(CX**2)+SX*ROTSIN+CZ*SZ*(CX**2)
     A22=ROTCOS*(SZ**2)+ROTCOS*((CZ*SX)**2)+(CX*CZ)**2
     A23=-SZ*CX*ROTSIN+CX*CZ*SX-CX*CZ*SX*ROTCOS
     A31=-ROTSIN*CX*CZ+SZ*SX*CX-SZ*SX*CX*ROTCOS
     A32=SZ*CX*ROTSIN+CZ*SX*CX-CZ*SX*ROTCOS*CX
     A33=SX**2+(CX**2)*ROTCOS
     DO 201 L=IFIRST,ILAST
     KK=MAXAT(L)
     DO 201 K=1,KK
     GROUP(L,K,1)=AGROUP(L,K,1)*A11+AGROUP(L,K,2)*A12+AGROUP(L,K,3)*A13
     GROUP(L,K,2)=AGROUP(L,K,1)*A21+AGROUP(L,K,2)*A22+AGROUP(L,K,3)*A23
     GROUP(L,K,3)=AGROUP(L,K,1)*A31+AGROUP(L,K,2)*A32+AGROUP(L,K,3)*A33
     DO 201 J=1,3
C    TRANSLATING BACK TO ORIGINAL POSITIONS
```

```
  201 GROUP(L,K,J)=GROUP(L,K,J)+GROUP(II,JJ,J)
      ANGLE(IZ)=0.0
      RETURN
      END
      SUBROUTINE SURFACE
      COMMON/BLOCKC/TOTALEN(8),TOTAL1(8),TOTAL2(8),TOTAL3(8)
      COMMON/BLOCKF/GROUP(9,20,6),MAXAT(9),IROTN(8,4)
      COMMON/BLOCKG/JUMP
      COMMON/BLOCKH/I
      COMMON/BLOCKJ/IZ
      JUMP=1
      F1=0.0
      F2=0.0
      F3=0.0
      IENDG=IROTN(IZ,1)
      IENDA=IROTN(IZ,2)
      JGRP1=IROTN(IZ,3)
      ISTOP=JGRP1-1
      MAXAT1=MAXAT(JGRP1)
      DO 164 JATOM1=1,MAXAT1
C     DECIDING WHAT THE ATOM IS
      IF (GROUP(JGRP1, JATOM1, 6)-2.) 100, 101, 102
C     ASSIGNING  THE 'EFFECTIVE NUMBER OF ELECTRONS'
  100 ELECN1=1.146
C     ASSIGNING THE VAN DER WAALS RADIUS
      VWRAD1=1.20
      GO TO 106
  101 ELECN1=6.876
      VWRAD1=1.70
      GO TO 106
  102 CONTINUE
      IF (GROUP(JGRP1,JATOM1,6)-5.) 103, 104, 105
  103 ELECN1=8.022
      VWRAD1=1.55
      GO TO 106
  104 ELECN1=9.168
      VWRAD1=1.52
      GO TO 106
  105 ELECN1=17.19
      VWRAD1=1.80
  106 CONTINUE
      DO 165 JPRIME=1,ISTOP
      JGRP2=ISTOP-JPRIME+1
      IF (JGRP2.EQ.IENDG.AND.JATOM1.EQ.1) GO TO 165
      MAXAT2=MAXAT(JGRP2)
      DO 64 JATOM2=1,MAXAT2
      IF (JGRP2.EQ.IENDG.AND.JATOM2.EQ.IENDA) GO TO 64
C     THIS DO LOOP CALCULATES INTERATOMIC DISTANCES, MINIMAL CONTACT
C     DISTANCES AND THE ENERGY OF INTERACTION OF PAIRS OF ATOMS, EACH
C     ONE ON A DIFFERENT GROUP.  IN ORDER TO DETERMINE MINIMUM CONTACT
C     PARAMETERS THE PROGRAM MUST RECOGNIZE THE NATURE OF THE ATOMS
C     INVOLVED.  IN ORDER TO DO THIS THE ATOMIC CODE NUMBERS ARE USED.
C     THEY ARE ----
C     1 - HYDROGEN            2 - CARBON            3 - NITROGEN
C     5 - OXYGEN              7 - PHOSPHORUS
```

```
      R IS THE INTERATOMIC DISTANCE
C     NOW DECODING ATOMIC NUMBERS IN ORDER TO CHECK FOR VIOLATIONS OF
C     CLOSEST ALLOWED INTERATOMIC DISTANCES AND TO ASSIGN INTERATOMIC
C     PARAMETERS.
C     VAN DER WAAL'S RADII ARE TAKEN FROM  A. BONDI, J. PHYS. CHEM.,
C     VOL. 68, 441, (1964).  MINIMUM CONTACT DISTANCES ARE TAKEN  TO BE
C     62 PER CENT OF THESE VALUES.
      IF (GROUP(JGRP2,JATOM2,6)-2.) 107, 108, 109
  107 ELECN2=1.146
      VWRAD2=1.20
      GO TO 113
  108 ELECN2=6.876
      VWRAD2=1.70
      GO TO 113
  109 CONTINUE
      IF (GROUP(JGRP2, JATOM2, 6)-5.) 110, 111, 112
  110 ELECN2=8.022
      VWRAD2=1.55
      GO TO 113
  111 ELECN2=9.168
      VWRAD2=1.52
      GO TO 113
  112 ELECN2=17.19
      VWRAD2=1.80
  113 VWDIAM=VWRAD1+VWRAD2
C     DMIN IS THE MINIMUM ALLOWED CONTACT
      DMIN=VWDIAM*.62
C     R IS THE INTERATOMIC DISTANCE
     0R=SQRT((GROUP(JGRP1,JATOM1,1)-GROUP(JGRP2,JATOM2,1))**2+(GROUP(JGR
     1P1,JATOM1,2)-GROUP(JGRP2,JATOM2,2))**2+(GROUP(JGRP1,JATOM1,3)-GROU
     2P(JGRP2,JATOM2,3))**2)
C     CHECKING TO SEE IF THE DISTANCE IS TOO SMALL
      IF (R-DMIN) 2,2,1
    2 JUMP=0
      GO TO 1000
   10A=365000.*GROUP(JGRP1,JATOM1,5)*GROUP(JGRP2,JATOM2,5)/(SQRT(GROUP(
     1JGRP1,JATOM1,5)/ELECN1)+SQRT(GROUP(JGRP2,JATOM2,5)/ELECN2))
      B=A*(VWDIAM**6)*.5
C     CALCULATION OF CHARGE-CHARGE ENERGY.  UNITS OF ENERGY ARE CALORIES
C     PER MOLE
   62 E1=(331816.*GROUP(JGRP1,JATOM1,4)*GROUP(JGRP2,JATOM2,4))/R
C     CALCULATION OF CHARGE-POLARIZABILITY ENERGY FOLLOWS.
     0E2=-165908.*((GROUP(JGRP1,JATOM1,4)**2)*GROUP(JGRP2,JATOM2,5)+
     1(GROUP(JGRP2,JATOM2,4)**2)*GROUP(JGRP1,JATOM1,5))/(R**4)
C     CALCULATION OF SIX-TWELVE POTENTIAL FOLLOWS.
      E3=B*(R**(-12))-A*(R**(-6))
      E1=E1+E1
      E2=E2+E2
      E3=E3+E3
   64 CONTINUE
  165 CONTINUE
  164 CONTINUE
C     ADDING THE RECENTLY CALCULATED ENERGY TO THAT CALCULATED FOR
C     OTHER BONDS EARLIER
      E=E1+E2+E3
```

```
      IF (IZ-1) 1000,3,4
    3 TOTALEN(1)=F
      TOTAL1(1)=F1
      TOTAL2(1)=F2
      TOTAL3(1)=F3
      GO TO 1000
    4 IZM1=IZ-1
      TOTALEN(IZ)=TOTALEN(IZM1)+F
      TOTAL1(IZ)=TOTAL1(IZM1)+F1
      TOTAL2(IZ)=TOTAL2(IZM1)+F2
      TOTAL3(IZ)=TOTAL3(IZM1)+F3
 1000 RETURN
      END
      SUBROUTINE EXHAUST
      COMMON/BLOCKC/TOTALEN(8),TOTAL1(8),TOTAL2(8),TOTAL3(8)
      COMMON/BLOCKD/ICOUNT,I1,I2,I3,I4,I5,I6,I7,I8
      COMMON/BLOCKE/DTHETA(8)
      COMMON/BLOCKJ/IZ
    1 FORMAT (8I2,4E17.10)
      WRITE (1,1) I1,I2,I3,I4,I5,I6,I7,I8,TOTALEN(8),TOTAL1(8),TOTAL2(8)
     1,TOTAL3(8)
      ICOUNT=ICOUNT+1
      CALL GRAPHIC
      RETURN
      END
      SUBROUTINE SETUP
      COMMON/SETUP/MAXPLOT,INDEX(20,2),BOTTOM(20,9,9)
    1 FORMAT (8I10)
      READ 1, MAXPLOT
      DO 105 IPLOT=1,MAXPLOT
      DO 104 II=1,9
      DO 104 JJ=1,9
  104 BOTTOM(IPLOT,II,JJ)=10.**300
  105 READ1, INDEX(IPLOT,1),INDEX(IPLOT,2)
      RETURN
      END
      SUBROUTINE GRAPHIC
      COMMON/BLOCKC/TOTALEN(8),TOTAL1(8),TOTAL2(8),TOTAL3(8)
      COMMON/BLOCKD/ICOUNT,I1,I2,I3,I4,I5,I6,I7,I8
      COMMON/SETUP/MAXPLOT,INDEX(20,2),BOTTOM(20,9,9)
      DO 19 IPLOT=1,MAXPLOT
      ICHOICE=INDEX(IPLOT,1)
      GO TO (1,2,3,4,5,6,7,8) ICHOICE
    1 M1=I1
      GO TO 9
    2 M1=I2
      GO TO 9
    3 M1=I3
      GO TO 9
    4 M1=I4
      GO TO 9
    5 M1=I5
      GO TO 9
    6 M1=I6
      GO TO 9
```

```
    7 M1=17
      GO TO 9
    8 M1=18
    9 ICHOICE=INDEX(IPLOT,2)
      GO TO (11,12,13,14,15,16,17,18) ICHOICE
   11 M2=11
      GO TO 19
   12 M2=12
      GO TO 19
   13 M2=13
      GO TO 19
   14 M2=14
      GO TO 19
   15 M2=15
      GO TO 19
   16 M2=16
      GO TO 19
   17 M2=17
      GO TO 19
   18 M2=18
   19 BOTTOM(IPLOT,M1,M2)=AMIN1(BOTTOM(IPLOT,M1,M2),TOTALEN(8))
      RETURN
      END
      SUBROUTINE VIEWER
      COMMON/SETUP/MAXPLOT,INDEX(20,2),BOTTOM(20,9,9)
    1 FORMAT (1H1          ////)
    2 FORMAT (57X,15H SURFACE NUMBER I2          /)
    3 FORMAT (62X, 6H BOND I2          //)
    4 FORMAT (15X, 9E13.5)
    5 FORMAT (4X, 5H BOND I2, 4X, 9E13.5)
      DO 100 IPLOT=1,MAXPLOT
      PRINT 1
      PRINT 2,IPLOT
      PRINT 3, INDEX(IPLOT,1)
      DO 101 J=1,4
  101 PRINT 4, (BOTTOM(IPLOT,IQ,J),IQ=1,9)
      PRINT 5, INDEX(IPLOT,2),(BOTTOM(IPLOT,IQ,5),IQ=1,9)
      DO 102 J=6,9
  102 PRINT 4, (BOTTOM(IPLOT,IQ,J),IQ=1,9)
  100 CONTINUE
      RETURN
      END
```

BRUTE input:

| | |
|---|---|
| 1st card | number of rigid groups, number of angles to be tried for each of the 8 possible bonds. |
| 2nd card | number of atoms in each group |
| 3rd card | 1st atom of the 1st group-label, X, Y, and Z coordinates, charge, polarizability and number to identify type of atom This format is repeated for all atoms in group 1, then group 2, etc. |
| Next card | group number at one end of the first bond (the smaller number), atom number at that end of the bond, group number at the other end of the bond, last group to be rotated, bond label, starting torsional angle |
| Next card | incremental angles (in degrees) to be tried for the 1st bond. The angles must sum to zero or the program exits. The previous two formats are repeated for the other 7 bonds. |
| Next card | number of 2 dimensional surfaces to be generated by Subroutine VIEWER |
| Next card | bond numbers for the 1st 2 dimensional surface. This format is repeated for other surfaces. |

```
 1 1 1 1 1 1
 2   +  2
2' ENDO RIBOSE O3'   -0.5334E+00-0.2808E+01-0.3387E+00-3.5030E-01+6.4000E-01 +5.0
2' ENDO RIBOSE C3'    0.5004E-01-0.1949E+01-0.1316E+01+2.6815E-01+9.3000E-01 +2.0
2' ENDO RIBOSE P      0.2628E+00-0.3259E+01 0.9869E+00+1.6390E-01+3.0000E+00 7.00
2' ENDO RIBOSE PO    -0.4258E+00-0.2662E+01 0.2184E+01-5.5160E-01+8.4000E-01 5.00
2' ENDO RIBOSE PO'    0.4314E+00-0.4732E+01 0.1003E+01-5.5160E-01+8.4000E-01 5.00
2' ENDO RIBOSE O5'    0.1596E+01-0.2482E+01 0.8274E+00-3.5030E-01+6.4000E-01 +5.0
2' ENDO RIBOSE C5'    0.1827E+01-0.1633E+01-0.3091E+00 2.0899E-01+9.3000E-01 +2.0
2' ENDO RIBOSE H5'    0.2719E+01-0.2060E+01-0.9638E+00+5.2950E-02+4.2000E-01 +1.0
 1 1 2 3    3'-P              +20.0000
+180.000-183.000
 2 1 3 3    P-O5'             +100.000
+180.000-183.000
 1 1 2 3    BONDLABEL         +0.00000
+0.00000
 1 1 2 3    BONDLABEL         +0.00000
+0.00000
 1 1 2 3    BONDLABEL         +0.00000
+0.00000
 1 1 2 3    BONDLABEL         +0.00000
+0.00000
 1 1 2 3    BONDLABEL         +0.00000
+0.00000
 1 1 2 3    BONDLABEL         +0.00000
+0.00000
          1
          1          2
```

```
      SUBROUTINE DERIV
      COMMON/BLOCK1/IROTN(8,4),MAXGRP,MAXAT(9),IZ,GROUP(9,18,6)
C     THESE HAVE THE SAME MEANING AS THEY DID IN NEWCLEO,BRUTE,FINDMIN,ETC.
      COMMON/BLOCK3/G1(8),G2(8),G3(8)
C     THESE ARE THE FIRST, SECOND AND THIRD DERIVATIVES (DV/DTHETA , ETC.)
      DIMENSION COORD(9,20,3), CYLIND(9,20,3)
      G1(IZ)=0.0
      G2(IZ)=0.0
      G3(IZ)=0.0
C     CONVERTING TO CYLINDRICAL COORDINATES
C     ALLIGNING THE BOND ALONG THE Y AXIS
C     TRANSLATING
      IEND1=IROTN(IZ,1)
      IEND2=IROTN(IZ,3)
      IATOM=IROTN(IZ,2)
      DO 1 I=1,3
    1 COORD(IEND2,1,I)=GROUP(IEND2,1,I)-GROUP(IEND1,IATOM,I)
C     ROTATING
      A=SQRT(COORD(IEND2,1,1)**2+COORD(IEND2,1,2)**2)
      B=SQRT(A**2+COORD(IEND2,1,3)**2)
      CZ=COORD(IEND2,1,2)/A
      SZ=COORD(IEND2,1,1)/A
      CX=A/B
      SX=COORD(IEND2,1,3)/B
C     ELEMENTS OF THE ROTATION MATRIX
      A31=-CZ*SX
      A32=-CZ*SX
      DO 3 JGRP=1,MAXGRP
      MAX=MAXAT(JGRP)
      DO 3 JATOM=1,MAX
      DO 2 I=1,3
    2 COORD(JGRP,JATOM,I)=GROUP(JGRP,JATOM,I)-GROUP(IEND1,IATOM,I)
      XCOORD=COORD(JGRP,JATOM,1)*CZ-COORD(JGRP,JATOM,2)*SZ
      ZCOORD=COORD(JGRP,JATOM,1)*A31+COORD(JGRP,JATOM,2)*A32+
     1COORD(JGRP,JATOM,3)*CX
C     CYLINDRICAL COORDINATE R
      CYLIND(JGRP,JATOM,1)=SQRT(XCOORD**2+ZCOORD**2)
C     CALCULATING COS(THETA)
      CYLIND(JGRP,JATOM,2)=ZCOORD/CYLIND(JGRP,JATOM,1)
C      CALCULATING SIN(THETA)
    3 CYLIND(JGRP,JATOM,3)=XCOORD/CYLIND(JGRP,JATOM,1)
      IEND3=IROTN(IZ,4)
      DO 4 JGRP1=1,MAXGRP
      IF (JGRP1.GE.IEND2.AND.JGRP1.LE.IEND3) GO TO 4
      MAXAT1=MAXAT(JGRP1)
      DO 7 JATOM1=1,MAXAT1
      IF (JGRP1.EQ.IEND1.AND.JATOM1.EQ.IATOM) GO TO 7
      IF (GROUP(JGRP1, JATOM1, 6)-2.) 100, 101, 102
  100 ELECN1=1.146
      VWRAD1=1.20
      GO TO 106
  101 ELECN1=6.676
      VWRAD1=1.70
      GO TO 106
  102 CONTINUE
```

```
      IF (GROUP(JGRP1,JATOM1,6)-5.) 103, 104, 105
  103 ELECN1=8.022
      VWRAD1=1.55
      GO TO 106
  104 ELECN1=9.168
      VWRAD1=1.52
      GO TO 106
  105 ELECN1=17.19
      VWRAD1=1.80
  106 CONTINUE
      DO 5 JGRP2=IEND2,IEND3
      MAXAT2=MAXAT(JGRP2)
      DO 6 JATOM2=1,MAXAT2
      IF (JATOM2.EQ.1.AND.JGRP2.EQ.IEND2) GO TO 6
C     THIS DO LOOP CALCULATES INTERATOMIC DISTANCES, MINIMAL CONTACT
C     DISTANCES AND THE ENERGY OF INTERACTION OF PAIRS OF ATOMS, EACH
C     ONE ON A DIFFERENT GROUP.  IN ORDER TO DETERMINE MINIMUM CONTACT
C     PARAMETERS THE PROGRAM MUST RECOGNIZE THE NATURE OF THE ATOMS
C     INVOLVED.  IN ORDER TO DO THIS THE ATOMIC CODE NUMBERS ARE USED.
C     THEY ARE ----
C     1 - HYDROGEN            2 - CARBON              3 - NITROGEN
C     5 - OXYGEN              7 - PHOSPHORUS.
C     R IS THE INTERATOMIC DISTANCE
C     NOW DECODING ATOMIC NUMBERS IN ORDER TO CHECK FOR VIOLATIONS OF
C     CLOSEST ALLOWED INTERATOMIC DISTANCES AND TO ASSIGN INTERATOMIC
C     PARAMETERS.
C     VAN DER WAAL'S RADII ARE TAKEN FROM  A. BONDI, J. PHYS. CHEM.,
C     VOL. 68, 441, (1964).  MINIMUM CONTACT DISTANCES ARE TAKEN  TO BE
C     70 PER CENT OF THESE VALUES.
      IF (GROUP(JGRP2,JATOM2,6)-2.) 107, 108, 109
  107 ELECN2=1.146
      VWRAD2=1.20
      GO TO 113
  108 ELECN2=6.876
      VWRAD2=1.70
      GO TO 113
  109 CONTINUE
      IF (GROUP(JGRP2, JATOM2, 6)-5.) 110, 111, 112
  110 ELECN2=8.022
      VWRAD2=1.55
      GO TO 113
  111 ELECN2=9.168
      VWRAD2=1.52
      GO TO 113
  112 ELECN2=17.19
      VWRAD2=1.80
  113 VWDIAM=VWRAD1+VWRAD2
     0R=SQRT((GROUP(JGRP1,JATOM1,1)-GROUP(JGRP2,JATOM2,1))**2+(GROUP(JGR
     1P1,JATOM1,2)-GROUP(JGRP2,JATOM2,2))**2+(GROUP(JGRP1,JATOM1,3)-GROU
     2P(JGRP2,JATOM2,3))**2)
     0A=365000.*GROUP(JGRP1,JATOM1,5)*GROUP(JGRP2,JATOM2,5)/(SQRT(GROUP(
     1JGRP1,JATOM1,5)/ELECN1)+SQRT(GROUP(JGRP2,JATOM2,5)/ELECN2))
      P=A*(VWDIAM**6)*.5
      D1E1=331816.*GROUP(JGRP1,JATOM1,4)*GROUP(JGRP2,JATOM2,4)/R**2
      D2E1=-2.*D1E1/R
```

```
      D3E1=-3.*D2E1/R
     0D1E2=-663632.*((GROUP(JGRP1,JATOM1,4)**2)*GROUP(JGRP2,JATOM2,5)+
     1(GROUP(JGRP2,JATOM2,4)**2)*GROUP(JGRP1,JATOM1,5))/(R**5)
      D2E2=-5.*D1E2/R
      D3E2=-6.*D2E2/R
      D1E6=-6.*A*(R**(-7))
      D2E6=-7.*D1E6/R
      D3E6=-8.*D2E6/R
      D1E12=12.*B*(R**(-13))
      D2E12=-13.*D2E12/R
      D3E12=-14.*D2E12/R
      0PTHETA=CYLIND(JGRP1,JATOM1,1)*CYLIND(JGRP2,JATOM2,1)*(CYLIND(JGRP1
     1,JATOM1,3)*CYLIND(JGRP2,JATOM2,2)-CYLIND(JGRP1,JATOM1,2)*CYLIND(JG
     2RP2,JATOM2,3))
      DR1=D1E1+D1E2+D1E6+D1E12
C     RUNNING SUM OF THE FIRST DERIVATIVE (WITH ANGLES IN RADIANS)
      G1(IZ)=DR1*PTHETA/R+G1(IZ)
C     PAIRWISE CONTRIBUTION TO D**2V/DR**2
      DR2=D2E1+D2E2+D2E6+D2E12
C     PAIRWISE CONTRIBUTION TO D**3V/DR**3
      DR3=D3E1+D3E2+D3E6+D3E12
     0PTHETA2=CYLIND(JGRP1,JATOM1,1)*CYLIND(JGRP2,JATOM2,1)*(CYLIND(JGR
     1P1,JATOM1,2)*CYLIND(JGRP2,JATOM2,2)+CYLIND(JGRP1,JATOM1,3)*CYLIND(
     2JGRP2,JATOM2,3))
      PTHETA3=-PTHETA
C     D**2R/DTHETA**2
      RTHETA2=PTHETA2/R-(PTHETA**2)*(R**(-3))
C     D**3R/DTHETA**3
     0RTHETA3=PTHETA3/R-3.*PTHETA*PTHETA2*(R**(-3))+3.*(PTHETA**3)*(R**(
     1-5))
C     RUNNING SUM OF THE SECOND DERIVATIVE (WITH ANGLES IN RADIANS)
      G2(IZ)=-DR1*RTHETA2-((PTHETA/R)**2)*DR2+G2(IZ)
C     RUNNING SUM OF THE THIRD DERIVATIVE (WITH ANGLES IN RADIANS)
     0G3(IZ)=DR1*RTHETA3+RTHETA2*PTHETA*DR2*3./R+((PTHETA/R)**3)*DR3
     1+G3(IZ)
    6 CONTINUE
    5 CONTINUE
    7 CONTINUE
    4 CONTINUE
C     CONVERTING FROM RADIANS TO DEGREES
      G1(IZ)=G1(IZ)*(3.141592654/180.)
      G2(IZ)=G2(IZ)*(3.141592654/180.)**2
      G3(IZ)=G3(IZ)*(3.141592654/180.)**3
      RETURN
      END
```

This report was prepared as an account of Government sponsored work. Neither the United States, nor the Commission, nor any person acting on behalf of the Commission:

A. Makes any warranty or representation, expressed or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

B. Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used in the above, "person acting on behalf of the Commission" includes any employee or contractor of the Commission, or employee of such contractor, to the extent that such employee or contractor of the Commission, or employee of such contractor prepares, disseminates, or provides access to, any information pursuant to his employment or contract with the Commission, or his employment with such contractor.