INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

EMPIRICAL COMPUTATION OF IR AND VCD SPECTRA OF RNA & SPECTROSCOPIC ANALYSIS OF THE TAT-TAR INTERACTION

By

Ting Xiang

A dissertation submitted to the Graduate Faculty in Chemistry in partial fulfillment of the requirement for the degree of Doctor of Philosophy, The City University of New York

1996

This manuscript has been read and accepted for the Graduate Faculty in Engineering in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

March 14, 1996
Date

Chair of Examining Committee

March 14, 1996
Date

Executive officer

Prof. Max Diem

Prof. John Lombardi

Prof. Gary Quigley

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

EMPIRICAL COMPUTATION OF IR AND VCD SPECTRA OF RNA & SPECTROSCOPIC ANALYSIS OF THE TAT-TAR INTERACTION

By

Ting Xiang

Adviser: Professor Dixie J. Goss

In the first part of this thesis, we present models of VCD and IR intensities computation which are based on the exciton approach. The models include degenerate extended coupled oscillator model (DECO), nondegenerate extended oscillator model (NECO) and pseudo-single strand approach (PSS). We also show the necessity to include vibrations other than the C=O stretching motion and improve our computation by adjusting the direction of dipole derivative with respect to the C=O bond direction. We report observed and computed IR and VCD spectra of a number of polyribonucleic acids in D_2O/buffer solution in the 1600-1750 cm^{-1} spectral region.

The experimental data are compared with results calculated using different levels of sophistication within the exciton approach.

In the second part of this thesis, presented the CD, VCD, stopped-flow and molecular modeling results of a short peptide, which contains the basic region of the HIV-1 Tat (Trans- activator) protein, interacting with ΔTAR, a shortened form of native TAR (Trans-activation response element) which is located at the 5' end of the HIV long terminal repeat. The results confirm the conformational changes of both molecules upon binding to each other. Measured and computed VCD spectra suggest a left handed helix structure for the peptide. Kinetic studies indicate a two step mechanism for the binding reaction of these two molecules with $k_1 = 0.7 \times 10^8$ $M^{-1}sec^{-1}$ and $k_2 = 69.4\ sec^{-1}$.

ACKNOWLEDGMENTS

For help in making this work possible, I would like to express my appreciation to Professor Dixie J. Goss, my mentor who led me into the biophysical chemistry field and taught me all I needed. Thank her for her guidance and support during the past four years. To Professor Max Diem for his patience and the knowledge he gave to me. To Professor Gary Quigley and Professor John Lombardi for taking time to read this manuscript and for giving me the benefit of their wisdom and insight. To my husband Yu Yang for his encouragement and help about this work. To my colleagues Dr. Miriam Gulotta, Dr. Louisa M. Balasta, Dr. Ping Xie, Dr. Ma Sha, Dr. Yahang Wang, Chin-chuan Wei, Tang Chen and Xiaosong Wang for the help I got from them.

SYMBOLS & ABBREVIATIONS

ε: molar absorbency

D_{0A}: dipole strength for a $0 \rightarrow A$ vibrational transition

R_{0A}: rotational strength for a $0 \rightarrow A$ vibrational transition

$\bar{\mu}_{0A}$: electric dipole transition moment for a $0 \rightarrow A$ vibrational transition

$\bar{m}_{0A}$: magnetic dipole transition moment for a $0 \rightarrow A$ vibrational transition

Ψ_0: wave-function of polymer at ground state

Ψ_A: wave-function of polymer at excited state

V: perturbation term in the Hamiltonian for a polymer

V_{ij}: interaction between group i and group j in a polymer

ϕ_i^A: wave-function of monomer i in the excited state

ϕ_i^0: wave-function of monomer i in the ground state

ψ_i: zero order wave-function

Ψ_k^0: corrected zero order wave-function

$\Delta\bar{\nu}_k$: perturbed energy

V: interaction matrix

$\bar{\mu}$: electric transition dipole moment of polymer

$\bar{m}$: magnetic transition dipole moment of polymer

V^P: interaction matrix for subgroup

PEM: photoelastic modulator

F_{ABS}: scale factor used in the calculated IR spectra

F_{VCD}: scale factor used in the calculated VCD spectra

DECO: degenerate extended coupled oscillator model

NECO: non-degenerate extended coupled oscillator model

PSS: pseudo single strand formalism

LTR: HIV-1 long terminal repeat

Tat: trans-activator

TAR: trans-activation response element

ϕ,φ : dihedral angles

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Part I

Empirical Computation of Infrared CD and Absorption Spectra of RNA

1. Introduction

Deoxyribonucleic acids (DNA) and ribonucleic acids (RNA) are chain like macromolecules. Their function is the storage and transfer of genetic information. These nucleic acids are essential components of all cells. The basic residues of the nucleic acids are deoxyribonucleotides and ribonucleotides. Each nucleotide is composed of a nitrogen-containing base (either of the pyrimidine or the purine type), a sugar (either a D-ribose or a 2-deoxy-D-ribose) and a phosphoric acid. The resonance structures of pyrimidine and purine bases are the sources of a high absorption of UV light with a absorption maximum in the region of 260-280nm. Nucleic acids are polycondensates consisting of nucleotides joined by 3' - 5' phosphodiester bonds. Three bases which follow one after the other constitute a base triplet. The linear order of the base triplet in DNA determine the sequence of the amino acid residues in the coded protein.

There are many kinds of chain structures observed in naturally occurring nucleic acids and prepared synthetic nucleic acids. The size of nucleic acid strands

varies enormously. The shortest natural RNAs are transfer RNAs (tRNA) with 75 - 84 residues. The longest RNAs are in the heterogeneous nuclear RNA (hnRNA) class and may contain up to 2×10^5 residues (Cantor & Schimmel, 1980). Most ribo and deoxyribo compounds prefer to form double helices with quite different geometry.

1.1 Double-Stranded Helix of Nucleic Acid

In 1953, James Watson and Francis Crick deduced the three dimensional structure of DNA and inferred its mechanism of replication. The important features of their model are: the two polynucleotide chains are coiled around a common axis, the chains run in opposite directions, the bases are on the inside of the helix, whereas the phosphate and deoxyribose units are on the outside, the two chains are held together by hydrogen bonds between pairs of bases, adenine is always paired with thymine, guanine is always paired with cytosine, the precise sequence of bases carries the genetic information (Saenger, 1984).

Many different helical geometries can be built around the fixed Watson - Crick base pairs. Helical parameters are the terms to express the different helical arrangement of polynucleotides (Saenger, 1984). The two bases in a base pair are not exactly coplanar. Dihedral angle between individual base planes is defined as

propeller twist θ_p. Base-pairs are not exactly perpendicular to the helix axis. Roll angle θ_R and twist angle θ_T describe the orientation of a base-pair with respect to the helix axis. The distance between the center of a base-pair to the helix axis is defined as the diameter of the helix. The pitch height, P, of a helix is the distance along the helix axis for one complete helix turn. n is the number of nucleotides per turn. h = p/n is the unit height. T = 360°/n is the unit twist of the rotation between one nucleotide and the nearby nucleotide. Sugar puckering has different modes too. The X-ray crystallographic studies on a variety of mononucleotides have found that the five-membered furanose ring ($C^{1'}$, $C^{2'}$, $C^{3'}$, $C^{4'}$, $O^{1'}$) is generally nonplanar. It can be puckered in a envelope form with four atoms tend to be in a plane and the fifth atom (which can be either $C^{2'}$or $C^{3'}$) is out of plane by 0.5 Å. This leads to four possible sugar ring conformations, $C^{2'}$ -endo, $C^{2'}$ -exo, $C^{3'}$ -endo and $C^{3'}$ -exo. In $C^{2'}$ -endo, the $C^{2'}$ lies above the furanose ring on the same side as the base and $C^{5'}$. In $C^{2'}$ -exo, the $C^{2'}$ lies below the ring. $C^{3'}$ -endo and $C^{3'}$ -exo are the same except the nonplanar atom is $C^{3'}$ instead of $C^{2'}$.

Generally, there are three types of helix: A-type ($C^{2'}$-endo, right handed), B-type ($C^{3'}$-endo, right handed), and Z-type (left handed). DNA can be A, B, or Z form, but RNA can only be A type because of the difference between RNA and DNA.

1.2 RNA Double Stranded Helix

The covalent structure of RNA differs from that of DNA in two respects. RNA contains ribose rather than deoxyribose. The other difference is that one of the four major bases in RNA is uracil (U) instead of thymine (T). X-ray diffraction and infrared dichroism studies on fibers and films of nucleic acids showed the different structure parameters which indicated the different configurations of deoxyribo- and ribonucleic acids (Bush, 1973). These differences are with respect to the bases, the phosphates and sugar residues. In B- form DNA, the $C^{1'}$, $C^{2'}$, $C^{4'}$, and $O^{1'}$ atoms of the deoxyfuranose ring lie almost in a perfect plane and only the $C^{3'}$ atom is a little removed from the plane. The plane of the bases lies perpendicular to the helix axis. The interval between the bases is 0.34 nm. Unit twist per residue is 36° so that after every 10 residues the structure repeats itself. Pitch height is about 3.4 nm. It is practically impossible to build this structure with RNA because the additional oxygen atom in the ribose would make too close of a Van der Waals contact (Stryer, 1988). However, RNA can adopt a modified double helical form in which the base pairs are tilted about 16° away from the perpendicular to the helix axis, a structure like A-DNA. The A-helix is wider and shorter than the B-helix. Sugar pucker adopts a $C^{3'}$ -endo form to give sufficient room for its $O^{2'}$ atom. RNA double helices display two major, structurally similar conformations depending on the salt concentration of the environment. At low ionic strength, the A-RNA double helix with 11-base-pairs per

helix turn predominates. If the salt concentration is raised to in excess of 20%, A-RNA is transformed into A'-RNA with 12-base-pairs per helix turn (Saenger, 1984). Depending on their biological function, native RNA species contain double-helical domains connected by single-stranded stretches and paired bases are more than unpaired (Hartman, 1973).

1.3 Physical Techniques Used in Secondary Structural Determination of Nucleic Acid

X-ray crystallographic method is the most definitive technique used to investigate the structures of nucleoside, nucleotide and nucleic acids. This technique provides information at different levels of detail depending on the molecular weight and organization of the material into crystalline or quasi-crystalline material. For small molecules (with molecular weight less than 2000 daltons) and some polynucleotide fragments or tRNA, which crystallize as single crystals, X-ray can yield precise structures and its resolution is a function of crystal quality. For some double-helical DNA and RNA, instead of being obtained in the form of single crystals suitable for X-ray analysis they crystallize as fibrillar aggregates which are arranged around the fibrous axis. Owing to the inherent disorder of molecules in the fiber and limited data, a systematic single crystal analysis is therefore impossible. The best one can do is to compare these data with the calculated data of models. By comparison

with the observed X-ray diffraction pattern, an incorrect model, which may not be rejected by calculation, can be indicated. Additional modification of the model, such as changing the backbone torsion angles, sugar puckering mode and minimization the potential energy, can be applied (Saenger, 1984). The practical limitations to the X-ray crystallographic technique are that growing crystals is a time-consuming and complicated procedure and it can be applied only to crystallizable systems. Nucleic acids in *vivo* exist in dilute aqueous solution, where structural differences from the crystalline state may exist. Thus it is important to have other physical techniques capable of giving structure information. Spectroscopy methods have made possible the direct study of nucleic acids in solutions.

Among spectroscopy methods, NMR yields the most detailed information about structure of the polynucleotide (Saenger, 1984). In the proton spectrum of an aqueous solution, base stacking in nucleic acids can be demonstrated by means of the ring-current effect in the aromatic residues which induces a environment sensitive chemical shift. If isoshielding curves are applied, an idea of the relative orientation of bases within a stacking system can be obtained. Hydrogen bond information can also be investigated conveniently. Since the proton which forms the H-bond is partially deshielded by the attraction to the proton acceptor, a shift to lower field of its resonance can be experienced. But measurements must be carried out in appropriate nonaqueous solvents because of rapid exchange with the protons in water. Using ^{13}C, ^{15}N, and ^{31}P techniques, details on molecular conformation can be obtained from spin-

spin coupling constants J which related with the torsion angle θ, $C^{2'}$-endo ⇔ $C^{3'}$-endo equilibrium constant $K_{eq.}$, and rotation about C-O bonds of phosphates and sugar ring, etc. Spin relaxation times and Nuclear Overhauser effect (NOE) can be used to analyzed the orientation of base relative to the sugar moiety. Combination of all the details of an NMR analysis can picture a structure characteristic of the molecule. The disadvantages of the NMR method are the time consuming running of the experiment, the complexity of the spectrum assignment and the high requirement of sample purity.

Ultraviolet (UV) spectra are sensitive to base-stacking effects in polynucleotides because the π-electron systems of the base units are the site of absorption and the stacking of the bases in the intact double helix decreases the absorption in the UV. By slow cooling and warming of samples, the helix and coil transformation can been observed. This behavior can also be used to determine the melting point of double stranded molecules.

IR spectra indicate vibrational states of molecules. An important practical advantage of infrared spectroscopy is that can be applied to most materials regardless of the physical properties of the samples. Infrared spectroscopy of molecules in solution has provided useful information about the structure and interaction of nucleic acids and their constituent molecules. The absorption bands arising from the in-plane vibrations of the C=O, C=C, and C=N groups of the bases appear in the 1450-1750 cm^{-1} region. The frequencies and intensities of these bands are sensitive to interactions such as hydrogen bonding and complex formation which involve the carbonyl oxygen

or the heterocyclic nitrogen. Hence, they can be used to measure base-pairing interactions, to identify the tautomeric forms of the bases and to determine other structural parameters. Because of the opacity of water to infrared radiation in the 1550 - 1750 cm^{-1} range, nucleic acids and polynucleotides are usually studied in D_2O solution so that the solvent is easily compensated. Chloroform or deuterochloroform are also widely used as solvents to study of the specificity of base pairing and the strength of hydrogen-bonding interactions. The Raman spectrum yields, in principle, the same information as the infrared spectrum. Infrared absorption occurs when a molecular vibration produces a change in the dipole moment, whereas Raman scattering occurs when the vibration modulates the polarizability. With laser excitation, the great advantage of Raman scattering for the investigation of the vibration of nucleic acids is that they can be studied in their natural conformation, that is in aqueous solution. Water is a very weak Raman scatterer and has very few Raman bands, which therefore disturb the the Raman spectrum less. Second, Raman spectrometers cover the entire vibrational frequency region whereas many infrared spectrometers can't. But chemical purity and homogeneity of the sample must be more carefully controlled for Raman than for infrared spectroscopy. Nucleic acids are complex polymers with a base unit containing more than 30 atoms. In general, the vibrational spectrum of a polymer with such a big unit is very complex. The theoretical calculation of the normal vibration is very helpful to get a reliable

assignment for each vibrational band and Raman spectral line. Neither infrared nor Raman spectroscopy can give a reliable result if it is used in dilute solutions.

One of the most fascinating aspects of nucleic acid structure is their helicity which leads to their interactions in a different manner with left versus right circularly polarized radiation, thus generating optical activity. Chiroptical methods are the only techniques capable of directly probing molecular optical activity.

Optical rotation dispersion (ORD) or electronic dichroism (CD) spectra measure the optical activity in the UV absorption band region. ORD and CD can be applied to determine asymmetry surrounding a chromophore and are especially useful for monitoring changes of local environment such as double stranded to single stranded transition and complex formation. In kinetics, CD spectroscopy can be used to following slow and medium rate conformational changes. Over the past decades, two forms of vibrational optical activity have become established. One is called Roman optical activity (ROA) which is the differential scattering of circularly polarized photos from chiral molecules. The other form is known as vibrational circular dichroism (VCD), the extension of electronic circular dichroism into the infrared vibrational region of the spectrum. In VCD, one measures the small difference in the absorption of the sample for left versus right circularly polarized incident infrared radiation. Further analytically useful chiroptical techniques are the magnetically induced activity (MORD, MCD) and fluorescent-detected circular dichroism (Hopper, 1983).

1.4 Application of VCD

Vibrational circular dichroism is a technique with great potential for stereochemical analysis. VCD is exceptionally sensitive to details of molecular stereo structure. Large differences can be seen between the spectra of similar molecules with different conformations or configurations. Over the last 20 years many such studies have been carried out with a purpose of relating the observed VCD spectra of various classes of molecules with their molecular structure. Many efforts have been oriented with a goal of actually calculating the VCD spectrum with a model.

VCD was first applied to investigate the chirality of small molecules. The first measurements were reported in 1973 (Hsu, et al., 1973) for crystals and then for solutions (Holzwarth, et al., 1974). With small molecules, studies have been focused on a more complete understanding of theoretical basis of VCD intensities and establishing relationships between observed VCD features and details of molecular structure and conformation. Due to difficulties in the description of VCD intensities, various types of approximate theoretical models have been used. Early theoretical models focused on the nuclear displacements associated with vibrational motion as the fundamental descriptor of VCD intensity. "Perturbed degenerate mode model" and "coupled oscillator model" describes the coupling of a pair of degenerate electric transition moments that are skewed dissymmetrically with respect to one another (Holzwarth, et al., 1972; Sugeta, et al., 1976). " The fixed partial charge model" describes VCD intensity in a normal mode as the sum of all pairwise contributions of

nuclear displacements (schellman, 1973). Later, the motion of electronic charge density has been taken into account as a distinct intensity contribution form, such as "localized molecular orbital model"(Nafie, et al., 1977), "atomic polar tensor model"(Freedman, et al., 1983)," dynamic polarization model" and "bond dipole model"(Polavarapu, 1983).

For polypeptides, VCD has been shown to be capable of distinguishing between the commonly acepted secondary structures *via* comparison of amide bands (Yasui, et al., 1987). The statistical analysis of the VCD of selected proteins may establish the relationship between quantitative VCD data in the amide I' region with secondary structure of proteins (Pancoska, et al., 1991). Also, VCD calculation and measurement has been used to study peptide β-turns observed in small cyclic peptides (Xie, et al., 1995).

In recent years, VCD has been applied to the elucidation of nucleic acid solution structure (Annamalai, et al., 1987, Gulotta, et al., 1989). In nucleic acids, the coupling of the carbonyl groups of the bases gives conformation-dependent VCD signals between 1550 cm^{-1} and 1750 cm^{-1}. These vibrational modes are in the achiral bases and dissymmetric coupling of these bases made VCD sensitive to the polymer geometry.

Due to the enormous size of the molecules being studied, *ab inito* quantum mechanical calculations to determine molecular vibrational frequencies and the appropriate electric and magnetic dipole derivative necessary to compute VCD and IR

absorption intensity are not useful. Semi-empirical calculations are not sufficiently accurate. The best method for VCD calculations of biological macromolecules is an empirical calculation, where the observed vibrational frequency and infrared intensity of a monomeric unit are used as input parameters, and VCD and absorption spectra of the polymer are calculated based on different levels of approximation within the exciton approach. This original calculation was referred to as the "extended coupled oscillator" model (Gulotta et al., 1989). These calculations assume that dipole-dipole coupling between the monomeric subunits is the only source of the induced optical activity. Baur and Keiderling (1993) have shown that the vibrational exciton approach is a valid approximation in the case of base vibrations in nucleic acids, which they call "weakly coupled transitions". In strongly coupled vibrations, such as the amide I vibrations in peptides, the interactions of the peptide linkages are partially due to dipole coupling, and partially due to through bond coupling.

The extended coupled oscillator approach is successful for guanine (G) and cytosine (C) containing poly- and oligonucleic acids in both A and B forms. In the case of adenine (A) and thymine (T) or uracil (U) containing nucleotides, the original calculations did not work out well, due to the fact that the interaction of two carbonyl groups in T or U appear too strongly in the calculations. Furthermore it was found that including vibrations other than the C=O stretching motion is necessary, since A does not contain any C=O groups, yet shows strong VCD in the 1600 - 1750 cm^{-1} region. The other bases, C, G, and T or U exhibit weak VCD couplets around 1630

cm^{-1}, in addition to the main VCD signals between 1650 and 1675 cm^{-1}. These lower frequency transitions have been ignored so far in the previous model calculation. In order to include these other vibrational modes , we need to consider the fact that their dipole transition moments could couple with those of C=O vibration. Thus, the formalism for VCD calculations of weakly interacting sets of nondegenerate oscillators was developed. In the case of nucleotides containing T or U, the single most important improvement of the computational results was obtained when the direction of the dipole derivative was adjusted with respect to the C=O bond direction. This correction is very important, particularly if the carbonyl stretching vibration couples with the ring vibrations of the base. However, these corrections require semi-empirical vibrational calculations to determine the amount of mixing of the vibrational coordinate. In Section 2, we will present the background and the equations for the understanding of model calculations.

2. Theoretical Background

Optical activity can result from inherently dissymmetric chromophore, such as a chiral molecule which has handedness. Optical activity can also be due to a molecule which is inherently symmetric, but dissymmetrically perturbed. For example, one can construct inherently dissymetric chromophores by coupling two inherently symmetric chromophores such that they are dissymmetrically disposed toward each other. Polynucleotides are optical active due to their dissymmetricly arranged achiral residues. Optical activity measurement includes optical rotatory dispersion (ORD), circular dichroism (CD) and vibrational circular dichroism (VCD) as well as magnetically induced activity (MORD, MCD) and fluorescence-detected circular dichroism. They are all used as tools to monitor structural changes of biopolymer molecules.

2.1 CD And VCD as a Monitor for Optical Active Molecule

To explain the principle of CD and VCD as a monitor for the optical activity, let us consider an unperturbed beam of plane polarized light (Figure 2-1). The light propagates along the Y direction. The electric vector **E** always lies on the YZ plane.

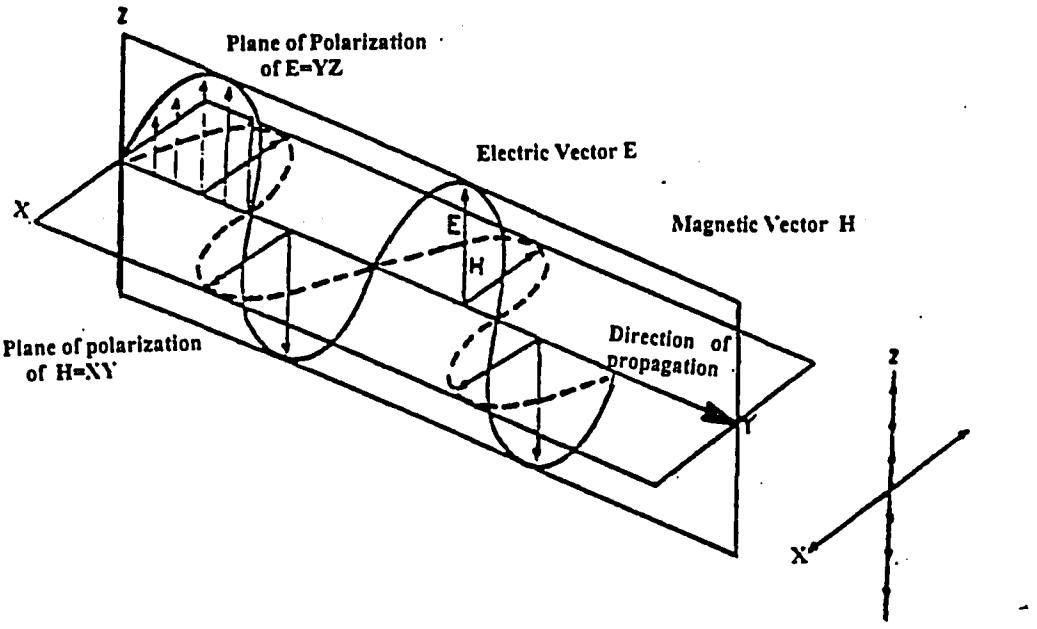


Figure 2-1. Electric E and magnetic H components of light polarized in YZ plane. When projected to XZ plane, E vector lies on Z axis.

When projected onto XZ plane, **E** vector lies entirely on Z axis. Equation (2-1)(Cantor, et al., 1980) describes the magnitude and orientation of **E** vector changing as a function of time.

$$\bar{E} = \bar{i} \cdot E_0 \cdot \sin\omega t \qquad (2-1)$$

We can also represent this light as the sum of two circularly polarized beams. They are oppositely oriented, left and right. Figure (2-2) is the right circularly

polarized component of plane polarized light. The E vector is no longer always on the YZ plane, but like a horizontal spiral staircase. When projected onto

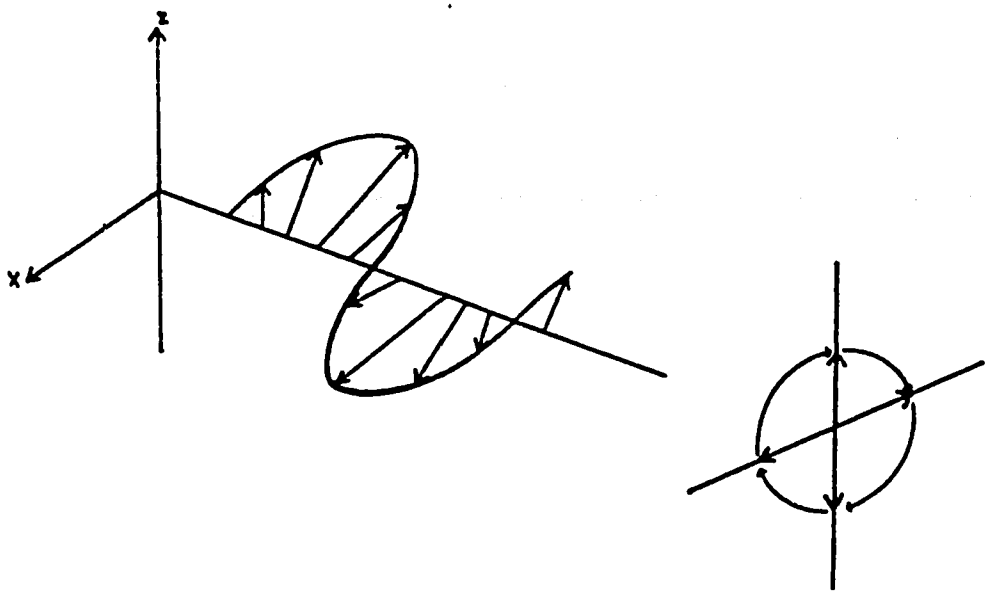


Figure 2-2. Right circularly polarized component of plane polarized light. When projected onto XZ plane, the resultant vector describes a circle.

the XZ plane, the vector describes a clockwise circle. For a left circularly polarized light, the circle is anti-clockwise. Equation (2-2) and (2-3) are for right and left circularly polarized light respectively.

$$\vec{E}_R = 1/2(\vec{i} \cdot E_0 \cdot \sin\omega t \quad + \vec{j} \cdot E_0 \cdot \cos\omega t) \qquad (2-2)$$

$$\vec{E}_L = 1/2(\vec{i} \cdot E_0 \cdot \sin\omega t \quad - \vec{j} \cdot E_0 \cdot \cos\omega t) \qquad (2-3)$$

The relationship among these three beams (right and left circularly polarized beams and plane polarized beam) can be verified easily. Let us consider the projection of two electric vectors $\mathbf{E_L}$ and $\mathbf{E_R}$ on the XZ plane at the same time. The XZ plane is perpendicular to the direction of propagation of the light wave. Because there is no interaction between light and medium, two circularly polarized beams have equal velocities and magnitudes. At this condition, the projections of these two electric vectors $\mathbf{E_L}$ and $\mathbf{E_R}$ would exhibit equal amplitudes and identical angles with the plane of polarization. Therefore, their sum, which is the resultant vector, would always lies in the plane of polarization. If we look from the Y direction, the projection of the resultant vector **E** would always move along the Z axis at any moment (Figure 2-3).

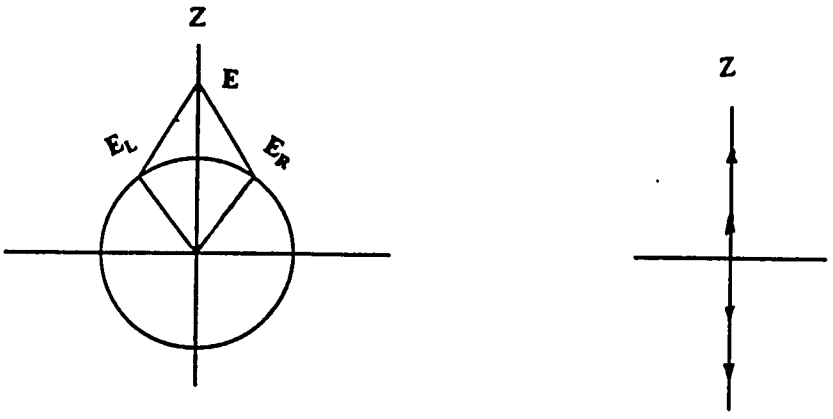


***Figure 2-3** Projection of right(R) and left(L) circularly polarized components of light projected onto XZ plane. $|E_L| = |E_R|$ and the angles between Z axis and L or R are always equal. Right side one is the resultant vector E lies along Z axis.*

We can also verify it by adding equation (2-2) and equation (2-3) together. The resultant equation is exactly the equation (2-1).

If a sample is an optical active material, it will interact with right and left circularly polarized beams unequally. As a consequence, it will unequally slow down the velocities of these two components of the light when they are passing though the sample ($v_L \neq v_R$). Since the refractive index is a measure of the velocity of light through the given material relative to its velocity in a vacuum ($n_L = v_o/v_L$, $n_R = v_o/v_R$) and thus $n_L \neq n_R$. Unequal velocities of these two beams will lead to unequal angles of the projected vectors on the XZ plane (Figure 2-4). The **E**, which is the resultant vector of the sum of $\mathbf{E}_L$ and $\mathbf{E}_R$, will still be in a plane, but that plane will exhibit an angle φ with respect to the plane of polarization of the incident light. The greater the difference of the interaction of the beams with the optically active sample, the larger the angle φ. Therefore, the angle is proportional to the difference between the refractive indexes for the right and left circularly polarized beams as $\varphi = (180^*l/\lambda)(n_L-n_R)_\lambda$. There, l is the length of the light path through the sample; φ, called optical rotation, is a function of the incident light wavelength λ. The curve describing the value of φ as a function of λ is the optical rotatory dispersion (ORD).

If the energy of the incident plane polarized beam corresponds to an absorbance band of the optically asymmetric sample, the sample will absorb the energy of these two components of incident light unequally ($\varepsilon_L \neq \varepsilon_R$), where ε is the

molar absorbency. $\varepsilon_L \neq \varepsilon_R$ leads to unequal magnitudes of vectors $\mathbf{E_L}$ and $\mathbf{E_R}$ as well as unequal amplitudes of their projections on the XZ plane. The resultant vector **E**

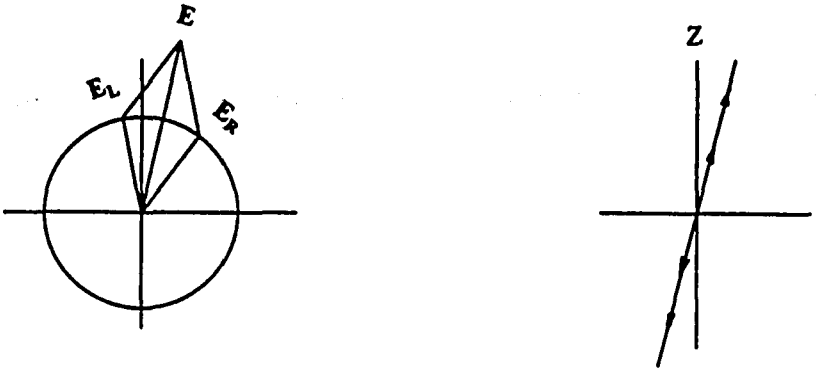


Figure 2-4. Projection of right(R) and left(L) circularly polarized components of light projected to the XZ plane, the resultant vector E lies in a new plane of polarization which is rotated at an angle with respect to plane YZ. $|E_L| = |E_R|$, *but* $\delta_L \neq \delta_R$

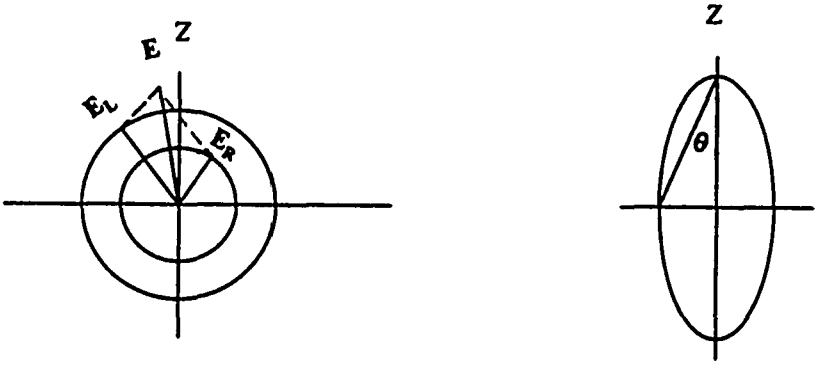


Figure 2-5. $\delta_L = \delta_R$, *but* $|E_L| \neq |E_R|$. *The resultant vector E lies in an ellipsoid.*

will no longer lie in a plane but an ellipsoid (Figure 2-5). Actually, both interaction and absorption will happen between samples and beams. Therefore, not only the velocities but also the magnitudes of the electric vectors of the right and left circularly polarized beams will be affected unequally. That is, not only will $n_L \neq n_R$, but also $\varepsilon_L \neq \varepsilon_R$. In considering again the same plane polarized light beam contains right and left circularly polarized components, their projection, $\mathbf{E}_L$ and $\mathbf{E}_R$, at any instant, not only make different angles with the original plane polarization (because $n_L \neq n_R$), but also have different magnitudes (because $\varepsilon_L \neq \varepsilon_R$). The locus traced by the resultant vector **E** of such a system will be an ellipsoid (Figure 2-6), whose major axis is inclined at an angle φ to the plane of the incident beam. It can be shown that the ratio of the lengths

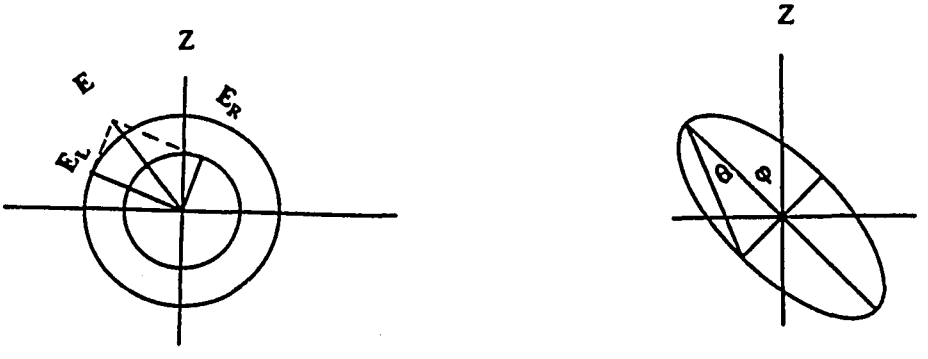


Figure 2-6. $\delta_L \neq \delta_R$ *and* $|E_L| \neq |E_R|$. *The resultant vector lies in an ellipsoid whose major axis is inclined at an angle* φ *to the plane of incident light.*

of the minor and the major axes of this ellipsoid, the tangent of the angle θ, is a measure of the difference between the energy absorbed by the sample from the right and the left circularly polarized beams at that wavelength. That is $\theta = (2.303*180/4\pi l)(\varepsilon_L - \varepsilon_R)$. There θ is a function of incident light wavelength λ. because ε_L and ε_R are function of incident light wavelength λ. The magnitude of θ as a function of incident light wavelength λ. is the circular dichroism. If the wavelength λ of the incident light is in the UV region, it corresponds to CD. If the wavelength λ of the incident light is in the IR region, it corresponds to VCD.

The actual shape, location, and magnitude of CD or VCD spectra of macromolecules are affected not only by the nature of the chromophore whose transition is contributing to the signal, but also by the three dimensional organization of these macromolecules. It will be explained in detail at the Section 2.2.

2.2 Computation Method

The intensity of a vibrational transition $0 \rightarrow A$ for a normal mode Q_a is proportional to the dipole strength D_{0A} defined as

$$D_{0A} = \left|\vec{\mu}_{0A}\right|^2 \qquad (2\text{-}4)$$

where $\vec{\mu}_{0A}$, the electric dipole transition moment, is related to the change in the molecular dipole moment due to the vibrational motion.

Vibrational circular dichroism arises from the interference of the electric dipole transition moment $\vec{\mu}_{0A}$ and the magnetic dipole transition moment $\vec{m}_{0A}$ and is proportional to the rotational strength R_{0A}.

$$R_{0A} = I_m(\vec{\mu}_{0A}) \cdot (\vec{m}_{0A}) \qquad (2\text{-}5)$$

I_m refers to the imaginary part of the scalar product. The rotational strength is positive or negative depending on the angle between the vectors.

In terms of transition matrix elements of the electric and magnetic dipole moment operators, the transition dipole moments are

$$\vec{\mu}_{0A} = \langle \Psi_A | \vec{\mu} | \Psi_0 \rangle = \langle \Psi_A | \sum_i e_i \vec{r}_i | \Psi_0 \rangle \qquad (2-6)$$

$$\vec{m}_{0A} = \langle \Psi_0 | \vec{m} | \Psi_A \rangle = \langle \Psi_0 | \sum_i \frac{e_i}{2m_i c} \vec{r}_i \times \vec{p}_i | \Psi_A \rangle \qquad (2-7)$$

for all the particles in the molecule with mass m_i, charge e_i, position r_i, linear momentum $\vec{p}_i$ and angular momentum $\vec{r}_i \times \vec{p}_i$. Ψ_0 and Ψ_A are the wavefunctions of polymer at ground and excited states.

The dipole and rotational strengths are related to the experimentally observed intensities by

$$\Delta\varepsilon = \frac{32\pi^3 \nu N_0}{6909hc} \sum_{A \neq 0} R_{0A} f(\nu - \nu_{0A}) \qquad (2-8)$$

$$\varepsilon(\gamma) = \frac{8\pi^3 \nu N_0}{6909hc} \sum_{A \neq 0} D_{0A} f(\nu - \nu_{0A}) \qquad (2-9)$$

ν: frequency in Hertz

N_0: Avogadro's number

h: Planck's constant

c: speed of light

$f(\nu-\nu_{0A})$ is a function which has the shape of an absorption band centered at frequency ν_{0A}. A gaussian function is often used for $f(\nu-\nu_{0A})$. $\Delta\varepsilon$ and ε are the sums of dipole and rotational strengths for all possible transitions in the wavelength region at which we did the measurement. From equation (2-8) and (2-9), one can see the VCD measurement $\Delta\varepsilon$ connects with the three dimensional organization of the molecule through wavefunctions. For a polymer, wave functions are very complicated. But using exciton theory, we can simplify the polymer wavefunctions.

2.2.1 Exciton Theory

The exciton model was proposed originally to explain the splitting of vibronic energy levels of formally degenerate states of solid molecules in a unit cell (Davydov, 1962). Later this model was adapted (Tinico, 1963; Bayley et al., 1969) to explain the optical activity of helical polymers. In this model, one assumes that the excited states of a polymer (or unit cell) composed of two or more monomers is delocalized over all the monomers. Thus, if one photon is absorbed into the excited state, it is impossible to determine which of the monomer units is actually excited. Rather, the excited state wave function is best described by a sum over all possible one-quantum excitations. This implies that the excitation is no longer localized on one of the oscillators, but is delocalized over the entire array of identical oscillators. This delocalized excitation is referred to as an "exciton". The dipolar coupling between the transition lifts their degeneracy; consequently, one observes as many discrete exciton energy levels as there are interacting dipoles.

The Hamiltonian for a system of N-monomers (or groups) was written as the sum of two types of terms. The first type contains terms pertains only to the individual groups; the second is the term which represents the potential energy of interaction between the groups.

$$H = \sum_{i=1}^{N} H_i + V \qquad (2-10)$$

$$= \text{zero-order Hamiltonian} + \text{perturbation}$$

$$V = \sum_{i} \sum_{j>i} V_{ij} \qquad (2-11)$$

H_i is the Hamiltonian for the group i in a vacuum. The solutions of H_i are the wavefunctions and energies of the isolated groups. V_{ij} is the interaction between group i and group j. We approximate V_{ij} by

$$V_{ij} = \vec{\mu}_{i0A} \cdot \vec{T}_{ij} \cdot \vec{\mu}_{j0A} \qquad (2-12)$$

$$\vec{T}_{ij} = \frac{1}{\vec{r}_{ij}^{\,3}}\left(1 - \frac{3\vec{r}_i \cdot \vec{r}_j}{\vec{r}_{ij}^{\,2}}\right) \qquad (2-13)$$

in (2 – 12):

$\vec{\mu}_{i0A} = \langle \phi_i^A | \vec{\mu}_i | \phi_i^0 \rangle$ is the electric dipole of the group i (2 - 14)

in (2 – 14):

$\vec{\mu}_i = \sum_s e\vec{r}_s'$ is the sum over all electrons and nucleus in group i (2 - 15)

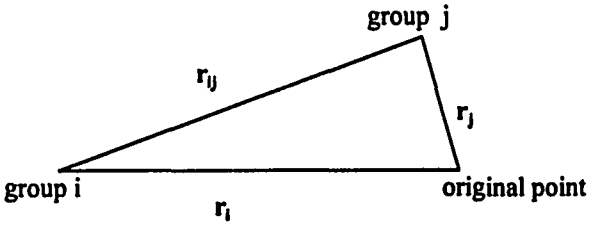


Figure 2-7. Relationship among the vectors in equation 2-13

Introduce (2-13) into (2-12):

$$V_{ij} = \frac{1}{\bar{r}_{ij}^{\,3}}[\vec{\mu}_{i0A} \cdot \vec{\mu}_{j0A} - 3(\vec{r}_{ij} \cdot \vec{\mu}_{i0A})(\vec{r}_{ij} \cdot \vec{\mu}_{j0A}) / \bar{r}_{ij}^{\,2}] \qquad (2\text{-}16)$$

The potential energies of the interaction between groups are treated as perturbations. With the ΣH_i as the zero-order Hamiltonian, we obtain the zero order wavefunction for the system as a simple products of group wave functions.

Therefore, the zero order wavefunction of ground state is:

$$\psi^0 = \phi_1^0 \phi_2^0 \cdots \phi_j^0 \cdots \phi_N^0 \qquad j = 1,2,\cdots,N \qquad (2\text{-}17)$$

(2-18) is the zero order wavefunction of an excited state with group i excited. Because the excited i can be anyone of the N groups, the functions are N-fold degenerate.

$$\psi_j^0 = \phi_1^0 \phi_2^0 \cdots \phi_j^A \cdots \phi_N^0 \qquad j = 1,2,\cdots,N \qquad (2\text{-}18)$$

The interaction among monomers will lead to the degenerate states splitting, resulting in corrected zero order wavefunctions which are the linear combination of zero order wavefunctions and no longer degenerate.

(2-19) is the corrected zero order wavefunctions of excited states:

$$\Psi_k^0 = \sum_{j=1}^{k} C_{jk}\psi_j^0 \qquad (2\text{-}19)$$

The C_{jk} and perturbed energies $\Delta\bar{\nu}_k$ (in vibrational frequency) can be conveniently obtained as eigenvectors and eigenvalues of the perturbation matrix $\{V_{ij}\}_{NXN}$. The elements of this symmetric matrix are equation (2-16). Next, we will derive the expressions for diople and rotational strengths using these wavefunctions.

2.2.2 D and R Expressions for a Dimer

Figure 2-8 (Holzwarth, et al., 1972) is a dimer of two groups, group **a** and group **b**. The index i designates for the particles in group **a** and the index j serves for particles in group **b**. $\bar{r}_i$ is the position vector of particle i with respect to a coordinate origin at the center of mass of group **a**. $\bar{y}_i$ is the position vector of particle i. $\bar{R}_a$ is a vector to the mass center of group **a** from an arbitrary origin. $\bar{r}_j$, $\bar{y}_j$ and $\bar{R}_b$ have the same meaning but corresponding to the group **b**. Equation (2-20) can be derived from Figure 2-8.

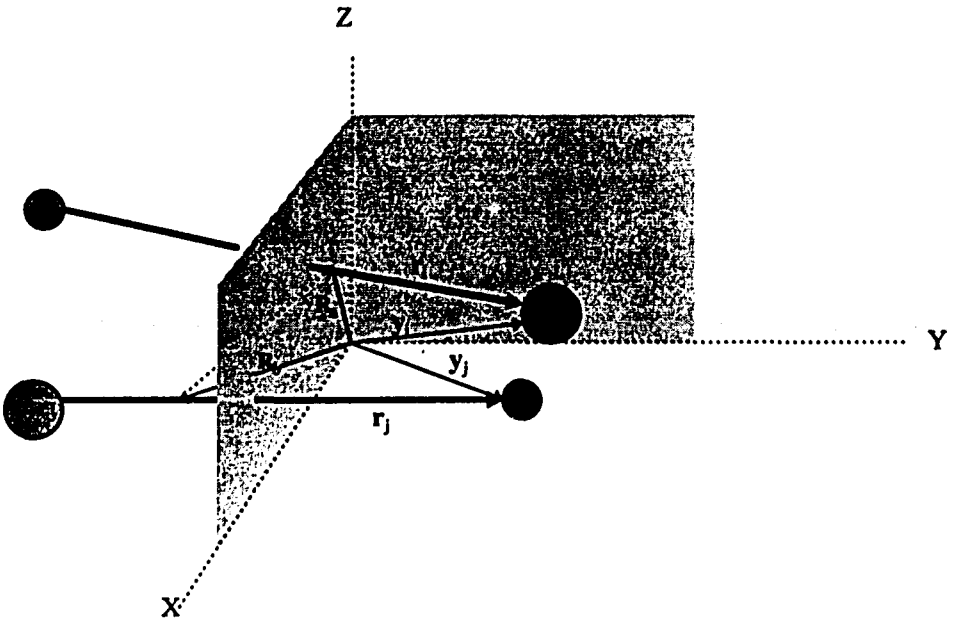


Figure 2-8. A dimer containing two identical diatomic molecules a and b.

$$\vec{y}_i = \vec{r}_i + \vec{R}_a$$
$$\vec{y}_j = \vec{r}_j + \vec{R}_b \qquad (2\text{-}20)$$

Because of the interaction between these two groups, the corrected zero order wavefunctions of this dimer are the linear combinations of two degenerate zero order wavefunctions $\phi_a^0\phi_b^A$ and $\phi_a^A\phi_b^0$.

$$\Psi_+^0 = C_1^+\phi_a^A\phi_b^0 + C_2^+\phi_a^0\phi_b^A \qquad (2-21)$$

$$\Psi_-^0 = C_1^-\phi_a^A\phi_b^0 + C_2^-\phi_a^0\phi_b^A \qquad (2-22)$$

$C_{1,2}^{+,-}$ and $E_{\pm}$ can be obtained by diagonalization of the matrix **V**.

$$\mathbf{V} = \begin{bmatrix} E_a^0 & V_{ab} \\ V_{ba} & E_b^0 \end{bmatrix}. \qquad (2\text{-}23)$$

here, $V_{ab} = V_{ba}$, E^0_a and E^0_b correspond to the fundamental frequencies of group **a** and group **b** without interactions. Based on equation (2-6), the electric transition dipole moment is:

$$\bar{\mu}^{\pm} = (\Psi_{\pm}^0 | \sum_i e_i \bar{y}_i + \sum_j e_j \bar{y}_j | \Psi^0) \qquad (2\text{-}24)$$

here $\Psi^0 = \phi_a^0 \, \phi_b^0$

$\sum_i$ is the sum over all particles in the group **a**.

$\sum_j$ is the sum over all particles in the group **b**.

Introducing (2-21) and (2-22) into (2-24), we get:

$$\bar{\mu}^{+} = (C_1^{+} \, \phi_a^{A} \, \phi_b^0 + C_2^{+} \, \phi_a^0 \, \phi_b^{A} \, | \sum_i e_i \bar{y}_i + \sum_j e_j \bar{y}_j | \, \phi_a^0 \, \phi_b^0) \qquad (2\text{-}25)$$

By use of (2-20) we obtain:

$$\bar{\mu}^{+} = \langle \, C_1^{+} \phi_a^{A} \phi_b^0 + C_2^{+} \phi_a^0 \phi_b^{A} \, | \, \sum_i e_i \bar{r}_i + \bar{R}_a \sum_i e_i + \sum_j e_j \bar{r}_j + \bar{R}_b \sum_j e_j | \, \phi_a^0 \, \phi_b^0 \, \rangle \qquad (2\text{-}26)$$

Because of $\int \phi_a^0 \phi_a^A d\tau = 0$, the terms $\vec{R}_a \sum_i e_i$ and $\vec{R}_b \sum_j e_j$ act as constants and contribute zero to the $\vec{\mu}^+$ integral.

$$\vec{\mu}^+ = \left\langle C_1^+ \phi_a^A \phi_b^0 \left| \sum_i e_i \vec{r}_i \right| \phi_a^0 \phi_b^0 \right\rangle + \left\langle C_1^+ \phi_a^A \phi_b^0 \left| \sum_j e_j \vec{r}_j \right| \phi_a^0 \phi_b^0 \right\rangle$$

$$+ \left\langle C_2^+ \phi_a^A \phi_b^0 \left| \sum_i e_i \vec{r}_i \right| \phi_a^0 \phi_b^0 \right\rangle + \left\langle C_2^+ \phi_a^A \phi_b^0 \left| \sum_j e_j \vec{r}_j \right| \phi_a^0 \phi_b^0 \right\rangle$$

$$= I_1 + I_2 + I_3 + I_4$$

$$I_1 = C_1^+ \int \phi_a^A \sum_i e_i \vec{r}_i \phi_a^0 d\tau \cdot \int \phi_b^0 \phi_b^0 d\tau$$

$$= C_1^+ \cdot \vec{\mu}_{a0A} \cdot 1$$

$$I_2 = C_1^+ \int \phi_b^0 \sum_j e_j \vec{r}_j \phi_b^0 d\tau \cdot \int \phi_a^A \phi_a^0 d\tau = 0$$

$I_3 = 0$ (similar as I_2)

$I_4 = C_2^+ \cdot \vec{\mu}_{b0A} \cdot 1$ (similar as I_1)

Therefore

$$\vec{\mu}^+ = C_1^+ \vec{\mu}_{a0A} + C_2^+ \vec{\mu}_{b0A} \qquad (2\text{-}27)$$

In the same way,

$$\vec{\mu}^- = C_1^- \vec{\mu}_{a0A} + C_2^- \vec{\mu}_{b0A} \qquad (2\text{-}28)$$

Here $\vec{\mu}_{a0A}$ or $\vec{\mu}_{b0A}$ is the electric transition dipole moment of the group **a** or group **b**. Its direction of dipole derivative is along the C=O bound. If we consider C=O stretching only, according to the equation (2-4), the magnitude of $\vec{\mu}_{a0A}$ or $\vec{\mu}_{b0A}$ is proportional to the monomer's (**a** or **b**) IR intensity of the C=O stretching mode.

Next, we will derive the expression for $\vec{m}^{\pm}$. According to equation (2-7):

$$\vec{m}^{\pm} = \left\langle \Psi^0 | \sum_i \frac{e_i}{2cm_i} \vec{y}_i \times \vec{p}_i + \sum_j \frac{e_j}{2cm_j} \vec{y}_j \times \vec{p}_j | \Psi_{\pm}^0 \right\rangle$$

$$\vec{y}_i = \vec{r}_i + \vec{R}_a \quad and \quad \vec{y}_j = \vec{r}_j + \vec{R}_b$$

$$\sum_i \frac{e_i}{2cm_i} \vec{y}_i \times \vec{p}_i + \sum_j \frac{e_j}{2cm_j} \vec{y}_j \times \vec{p}_j$$

$$= \frac{1}{2c} \left(\sum_i \frac{e_i \vec{r}_i \times \vec{p}_i}{m_i} + \sum_i \frac{e_i \vec{R}_a \times \vec{p}_i}{m_i} + \sum_j \frac{e_j \vec{r}_j \times \vec{p}_j}{m_j} + \sum_j \frac{e_j \vec{R}_b \times \vec{p}_j}{m_j} \right) \quad (2-29)$$

Use equations (2-21), (2-22) and (2-29) to express $\vec{m}^{\pm}$:

$$\bar{m}^{\pm} = \langle\ \Psi^0 | \frac{1}{2c}\left(\sum_i \frac{e_i \vec{r}_i \times \vec{p}_i}{m_i} + \sum_i \frac{e_i \vec{R}_a \times \vec{p}_i}{m_i} + \sum_j \frac{e_j \vec{r}_j \times \vec{p}_j}{m_j} + \sum_j \frac{e_j \vec{R}_b \times \vec{p}_j}{m_j}\right) | \Psi^0_{\pm}\ \rangle$$

$$\Psi^0 = \phi^0_a\, \phi^0_b$$

$$\Psi^0_{\pm} = C^{\pm}_1 \phi^{A}_a \phi^0_b + C^{\pm}_2 \phi^0_a \phi^{A}_b$$

$$\bar{m}^{\pm} = \langle\ \phi^0_a\, \phi^0_b | \frac{1}{2c}\left(\sum_i \frac{e_i \vec{r}_i \times \vec{p}_i}{m_i} + \sum_i \frac{e_i \vec{R}_a \times \vec{p}_i}{m_i} + \right.$$

$$\left. \sum_j \frac{e_j \vec{r}_j \times \vec{p}_j}{m_j} + \sum_j \frac{e_j \vec{R}_b \times \vec{p}_j}{m_j}\right) | C^{\pm}_1 \phi^{A}_a \phi^0_b + C^{\pm}_2 \phi^0_a \phi^{A}_b\ \rangle$$

$$\bar{m}^{\pm} = (I_1 + I_2 + I_3 + I_4)$$

here $\quad I_1 = \frac{1}{2c}\left[\int \phi^0_a \phi^0_b \sum_j \frac{e_i \vec{r}_i \times \vec{p}_i}{m_i} (C^{\pm}_1 \phi^{A}_a \phi^0_b + C^{\pm}_2 \phi^0_a \phi^{A}_b) d\tau\right]$

$$= \frac{C^{\pm}_1}{2c} \int \phi^0_a \phi^0_b \sum_i \frac{e_i \vec{r}_i \times \vec{p}_i}{m_i} \phi^{A}_a \phi^0_b d\tau + 0$$

$$= C^{\pm}_1 \int \phi^0_a \phi^0_b \cdot \int \phi^0_a \sum_i \frac{e_i \vec{r}_i \times \vec{p}_i}{2cm_i} \phi^{A}_a d\tau$$

$$= C^{\pm}_1 \cdot \int \phi^0_a \vec{m}_a \phi^{A}_a d\tau$$

$$= C^{\pm}_1\, \bar{m}_{a0A}$$

is the magnetic dipole moment of monomer **a**. For an achiral molecule, $\bar{m}_{a0A} = 0$. In the same way, $\bar{m}_{b0A} = 0$. Therefore, terms I_1 and I_3 contribute zero to $\bar{m}^{\pm}$. I_2 generates a term in $\bar{m}^{\pm}$ as:

$$I_2 = \frac{1}{2c}[\bar{R}_a \times \langle \phi_a^0 \phi_b^0 | \sum_i \frac{e_i \bar{p}_i}{m_i} | C_1^{\pm} \phi_a^A \phi_b^0 + C_2^{\pm} \phi_a^0 \phi_b^A \rangle] \qquad (2\text{-}30)$$

Since: $\langle \alpha | e_i \bar{p}_i / g_i | \beta \rangle = \langle \alpha | e_i \bar{r}_i | \beta \rangle \cdot (-2\pi i \nu_a^0)$

and $\langle \alpha | \bar{m} | \beta \rangle = -\langle \beta | \bar{m} | \alpha \rangle$ (Moffit, 1956)

Therefore, the equation (2-30) can be written as

$$I_2 = \frac{2\pi i \nu_a^0}{2c} \bar{R}_a \times [\langle C_1^{\pm} \phi_a^A \phi_b^0 + C_2^{\pm} \phi_a^0 \phi_b^A | \sum_i e_i \bar{r}_i | \phi_a^0 \phi_b^0 \rangle]$$

$$= \frac{2\pi i \nu_a^0}{2c} \bar{R}_a \times \bar{\mu}_{a0A} \cdot C_1^{\pm} \qquad (2\text{-}31)$$

In the same way I_4 generates a term in $\bar{m}^{\pm}$ as :

$$I_4 = \frac{2\pi i \nu_b^0}{2c} \bar{R}_b \times \bar{\mu}_{b0A} \cdot C_2^{\pm} \qquad (2\text{-}32)$$

Combining (2-31) and (2-32), we get following:

$$\vec{m}^{\pm} = \frac{\pi i}{c}[\nu_a^0 C_1^{\pm} \vec{R}_a \times \vec{\mu}_{a0A} + \nu_b^0 C_2^{\pm} \vec{R}_b \times \vec{\mu}_{b0A}] \qquad (2-33)$$

According to the equation (2-5):

$$R^{\pm} = I_m[\vec{\mu}^{\pm} \cdot \vec{m}^{\pm}]$$

$$= \frac{\pi}{c}[(C_1^{\pm} \vec{\mu}_{a0A} + C_2^{\pm} \vec{\mu}_{b0A}) \cdot (\nu_a^0 C_1^{\pm} \vec{R}_a \times \vec{\mu}_{a0A} + \nu_b^0 C_2^{\pm} \vec{R}_b \times \vec{\mu}_{b0A})]$$

$$= \frac{\pi}{c}[C_1^{\pm^2} \nu_a^0 \vec{\mu}_{a0A} \cdot (\vec{R}_a \times \vec{\mu}_{a0A}) + C_1^{\pm} C_2^{\pm} \nu_b^0 \vec{\mu}_{a0A} \cdot (\vec{R}_b \times \vec{\mu}_{b0A})$$

$$+ C_1^{\pm} C_2^{\pm} \nu_a^0 \vec{\mu}_{b0A} \cdot (\vec{R}_a \times \vec{\mu}_{a0A}) + C_2^{\pm^2} \nu_b^0 \vec{\mu}_{b0A} \cdot (\vec{R}_b \times \vec{\mu}_{b0A})]$$

$\because \quad \vec{a} \cdot (\vec{b} \times \vec{c}) = \vec{b} \cdot (\vec{c} \times \vec{a}) = \vec{c} \cdot (\vec{a} \times \vec{b})$

$$\therefore R^{\pm} = \frac{\pi}{c}[C_1^{\pm^2} \nu_a^0 \vec{R}_a \cdot (\vec{\mu}_{a0A} \times \vec{\mu}_{a0A}) + C_1^{\pm} C_2^{\pm} \nu_a^0 \vec{R}_a \cdot (\vec{\mu}_{a0A} \times \vec{\mu}_{b0A})$$

$$+ C_1^{\pm} C_2^{\pm} \nu_b^0 \vec{R}_b \cdot (\vec{\mu}_{b0A} \times \vec{\mu}_{a0A}) + C_2^{\pm^2} \nu_b^0 \vec{R}_b \cdot (\vec{\mu}_{b0A} . \times \vec{\mu}_{b0A})]$$

$\because \quad \vec{\mu}_{a0A} \times \vec{\mu}_{a0A} = \vec{\mu}_{b0A} . \times \vec{\mu}_{b0A} = 0$

$$\therefore R^{\pm} = \frac{\pi}{c}[C_1^{\pm} C_2^{\pm} \nu_b^0 \vec{R}_b \cdot (\vec{\mu}_{b0A} \times \vec{\mu}_{a0A}) + C_1^{\pm} C_2^{\pm} \nu_a^0 \vec{R}_a \cdot (\vec{\mu}_{a0A} \times \vec{\mu}_{b0A})]$$

$$= \frac{\pi}{c} C_1^{\pm} C_2^{\pm} (\nu_a^0 \vec{R}_a - \nu_b^0 \vec{R}_b) \cdot (\vec{\mu}_{a0A} \times \vec{\mu}_{b0A}) \qquad (2-34)$$

According to the equation (2-4)

$$D^{\pm} = (C_1^{\pm}\ \vec{\mu}_{a0A} + C_2^{\pm}\vec{\mu}_{b0A})\cdot(C_1^{\pm}\vec{\mu}_{a0A} + C_2^{\pm}\vec{\mu}_{b0A})$$

$$= C_1^{\pm^2}\vec{\mu}_{a0A}\cdot\vec{\mu}_{a0A} + C_2^{\pm^2}\vec{\mu}_{b0A}\cdot\vec{\mu}_{b0A} + 2C_1^{\pm}C_2^{\pm}(\vec{\mu}_{a0A}\cdot\vec{\mu}_{b0A}) \qquad (2-35)$$

2.2.3 NECO Model

If we extend equations (2-23),(2-34) and (2-35) to a N-polymer, we get the expression of a non-degenerate extended coupled oscillator (NECO) model:

$$V = \begin{bmatrix} E_1^0 & V_{12} & V_{13} & \cdots \\ V_{21} & E_2^0 & V_{23} & \cdots \\ V_{31} & V_{32} & E_3^0 & \cdots \\ \cdots & \cdots & \cdots & \end{bmatrix} \qquad (2-36)$$

$$R_k = -\frac{\pi}{c}\sum_{l=1}^{N}\sum_{m>l}^{N} C_m^k C_l^k(\nu_l^0\vec{R}_l - \nu_m^0\vec{R}_m)\cdot(\vec{\mu}_{l0A}\times\vec{\mu}_{m0A}) \qquad k = 1.2,\cdots N \quad (2-37)$$

$$D_k = \sum_{l=1}^{N}\sum_{m=1}^{N} C_l^k C_m^k(\vec{\mu}_{l0A}\cdot\vec{\mu}_{m0A}) \qquad k = 1.2,\cdots N \quad (2-38)$$

From the above equations, we can compute VCD and IR spectra if we know the followings:

$\vec{E}_l$ or ν_l^0, bserved vibrational frequency of the monomeric unit l.

$\vec{R}_l$, the vector connecting the center of mass of the monomeric unit l with the arbitrary origin.

$\vec{\mu}_{l0A}$, the electric dipole transition moment which is related to the change in the molecular dipole moment of the monomeric unit l due to the vibrational motion.

For example, if we consider a fundamental normal mode of C=O stretching only, $\vec{\mu}_{l0A}$'s magnitude is obtained by numerically integrating the experimental IR absorption spectrum of the monomer l over the carbonyl region. It's direction is the stretching direction which is along C=O bonds.

The empirical calculations of the VCD and IR spectrum for biological macromolecules are based on different levels of approximations with the exciton approach where the observed vibrational frequency and infrared absorption intensity of a monomeric unit are used as input parameters. The degenerate extended coupled oscillator (DECO) model give us a simple way to do the calculation.

2.2.4 DECO Model

With the assumption of that the N-polymer contains a single type of oscillator and the end effect is neglected, equations (2-37) and (2-38) are reduced to :

$$D_k = \sum_{i=1}^{N}\sum_{j=1}^{N} C_i^k C_j^k (\vec{\mu}_{i0A} \cdot \vec{\mu}_{j0A}) \qquad k = 1.2, \cdots N \quad (2-39)$$

$$R_k = -\frac{\pi \nu^0}{c}\sum_{i=1}^{N}\sum_{j>i}^{N} C_i^k C_j^k [\vec{T}_{ij} \cdot (\vec{\mu}_{i0A} \times \vec{\mu}_{j0A})] \qquad k = 1.2, \cdots N \quad (2-40)$$

Equations (2-39) and (2-40) are the rotational strengths R and dipole strengths D and hence the VCD intensities and infrared absorption intensities, for the kth exciton component. Where c is the velocity of light, and the C_i^k are the eigenvector components of the (dipole-dipole) interaction matrix V

$$V = \begin{bmatrix} 0 & V_{12} & V_{13} & \cdots \\ V_{21} & 0 & V_{23} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & 0 \end{bmatrix}_{N\times N} \qquad (2-41)$$

$$V_{ij} = \frac{\vec{\mu}_{i0A} \cdot \vec{\mu}_{j0A}}{\left|\vec{T}_{ij}\right|^3} - \frac{3(\vec{\mu}_{i0A} \cdot \vec{T}_{ij}) \cdot (\vec{\mu}_{j0A} \cdot \vec{T}_{ij})}{\left|\vec{T}_{ij}\right|^5} \qquad (2-42)$$

$$\vec{T}_{ij} = \vec{X}_j - \vec{X}_i \qquad (2-43)$$

$\vec{T}_{ij}$ is the distance vector between dipoles $\vec{\mu}_{i0A}$ and $\vec{\mu}_{j0A}$. $\vec{X}_j$ and $\vec{X}_i$ are the coordinate of the center of mass of oscillator i and j. ν^0 is the frequency (in wavenumbers) of the unperturbed transition. **V** is a symmetric matrix with zero along the diagonal and with all dipole interactions of the molecules at their appropriate places. The perturbed energy (in wavenumbers) for each of the kth exciton levels is given by:

$$\nu_k = \nu^0 + \sum\sum C_i^k C_j^k V_{ij} / h \qquad (2\text{-}44)$$

where the second term is the eigenvalues of the interaction matrix **V**. Both rotational and dipole strengths are converted to simulated spectra by overlaying (Gaussian, Lorenzian, or mixed) band profiles with suitable parameters.

In the dimeric case, the perturbation matrix **V** is given by:

$$V = \begin{bmatrix} 0 & V_{12} \\ V_{21} & 0 \end{bmatrix} \qquad (2\text{-}45)$$

(with $V_{12} = V_{21}$), which has eigenvectors with elements $\pm 1/2\sqrt{2}$. When these values are substituted into Equation (2-40), the well known case of the dimeric coupled oscillator is obtained (Gulotta et al., 1989) which predicts the rotational strengths R

for the symmetric and antisymmetric combination states of two interacting vibrations according to

$$R^{\pm} = \mp(\pi\nu^0 / 2c)\vec{T}_{12} \cdot \vec{\mu}_{10A} \times \vec{\mu}_{20A} \qquad (2\text{-}46)$$

2.2.5 PSS Calculation

The "Pseudo Single-Strand" formalism is a simplification of the NECO approach. It was first developed by Rhodes (1961) who treated double stranded helices as a single strand of degenerate dimers. In our approach, we consider a polymer of S identical subunits and T degenerate or non-degenerate transitions in each subunit, which may be a set of base pairs. The total number of oscillators in the polymer is N, as before. Thus, the number of subunits S is given by S = N / T and the Hamiltonian for this system is:

$$H = \sum_{k=1}^{S} H_k + V \qquad (2\text{-}47)$$

Where H_k is the Hamiltonia for the kth subunit and V is the interaction among S subunits.

To derive the dipole strength and rotational strength expressions, first we consider the interaction among the monomers in the kth subunit, $1 \leq k \leq S$. From equations (2-36) we write the interaction matrix for this subunit as:

$$\mathbf{V} = \begin{bmatrix} E_1^{0k} & V_{12}^k & \cdots & \cdots \\ V_{21}^k & E_2^{0k} & & \\ \cdots & \cdots & \ddots & \\ & & & E_T^{0k} \end{bmatrix} \qquad (2\text{-}48)$$

with eigenvalues ε^P_k and eigenvectors C^P_{ik} (where P = 1,2,... T). The zero order wavefunctions of the excited states of this subunit are:

$$\Psi_P^{0k} = \sum_{i=1}^{T} C_{ik}^P \psi_i^0 \qquad P = 1, 2, \ldots T \quad (2\text{-}49)$$

Next, we consider the interaction among S subunits. Corrected zero order wavefunctions of the polymer are the linear combinations of S × T wavefunctions expressed by equation (2-49). If we neglect the end effect, all the H_k are the same. Since H and H_k (k = 1, 2, ... S) commute, only the functions with the same eigenvalues can be mixed together. To see this more clearly, we write down the zero order wavefunctions (according to equation 2-49) of the excited states of all the subunits in groups. At each group, all the wavefunctions will have the same eigenvalue. For example, Ψ_p^{01} Ψ_p^{02} Ψ_p^{03} ... Ψ_p^{0k} ... Ψ_p^{0S} ($1 < k < S$) is a group which

has the wavefunctions of the Pth exciton in every subunit and has eigenvalues ε_1^P ε_2^P ε_3^P ... ε_k^P ... ε_S^P. Because we assume that S subunits are identical, $\varepsilon_i^P = \varepsilon_j^P$ and therefore, the subscript of ε can be omitted.

Ψ_1^{01} Ψ_1^{02} Ψ_1^{03} ... Ψ_1^{0k} ... Ψ_1^{0S} with eigenvalue ε^1

Ψ_2^{01} Ψ_2^{02} Ψ_2^{03} ... Ψ_2^{0k} ... Ψ_2^{0S} with eigenvalue ε^2

.....

Ψ_P^{01} Ψ_P^{02} Ψ_P^{03} ... Ψ_P^{0k} ... Ψ_P^{0S} with eigenvalue ε^P

.....

Ψ_T^{01} Ψ_T^{02} Ψ_T^{03} ... Ψ_T^{0k} ... Ψ_T^{0S} with eigenvalue ε^T

The interaction matrix has a form :

$$\mathbf{V} = \begin{bmatrix} V^1 & & & 0 \\ & V^2 & & \\ & & \ddots & \\ 0 & & & V^T \end{bmatrix} . \qquad (2\text{-}50)$$

Where V^P (P = 1, 2, ... T) is a S × S submatrix with the matrix elements (represents the interaction between two subunits n and m) of the form :

$$V_{n,m}^{P} = \sum_{i=1}^{T}\sum_{j=1}^{T} C_{in}^{P} C_{jm}^{P} V_{ni,mj} \qquad P = 1,2,\cdots T \qquad (2\text{-}51)$$

where $V_{ni,mj}$ is the interaction term of the ith oscillator in the subunit n with the jth oscillator in the subunit m. The expression for $V_{ni,mj}$ is as same as equation (2-42). With the assumption of S identical subunits, the C^P_{in} and C^P_{jm} can be written as C^P_i and C^P_j. By diagonalization of each submatrix $[\ V^P\ ]_{SXS}$, we get eigenvalues $\Delta E_k^{\ P}$ and eigenvectors ($A_{1k}^{\ P}$, $A_{2k}^{\ P}$, ... $A_{Sk}^{\ P}$), here P=1,2, ... T and k=1,2, ... S.

The complete set of transition frequencies and rotational and dipole strengths are given by:

$$E_k^{\ P} = \Delta E_k^{\ P} + \varepsilon^P \qquad P = 1,2, \ldots T \qquad (2\text{-}52)$$

ε^P came from the eigenvalues of matrix (2-48) with k omitted because we assumed S identical subunits.

$$D_k^P = \sum_{i=1}^{T}\sum_{j=1}^{T}\sum_{m=1}^{S}\sum_{n=1}^{S} C_i^P C_j^P A_{mk}^P A_{nk}^P \vec{\mu}_{mi} \cdot \vec{\mu}_{nj} \qquad k = 1,2,\cdots S \qquad (2-53)$$

$$R_k^P = \frac{\pi}{c}\{\sum_{i=1}^{T}\sum_{j=1}^{T}\sum_{m=1}^{S}\sum_{n=1}^{S} C_i^P C_j^P A_{mk}^P A_{nk}^P \vec{R}_{nj} \cdot (\vec{\mu}_{nj} \times \vec{\mu}_{mi})\} \quad P = 1,2,\cdots T \qquad (2-54)$$

The advantage of this method is that the largest matrix which needs to be diagonalized is S×S, where in NECO and DECO calculations , the matrix size is $(TxS)^2$.

For all computational approaches, the resulting dipole and rotational strength (in esu^2cm^2), were converted to extinction coefficient ε or Δε (in L mol^{-1} cm^{-1}) *via* the approximations:

$$D \approx 9.2 \times 10^{-39} \pi \varepsilon_{max} w / \nu^0 \qquad (2\text{-}55)$$

$$R \approx 2.3 \times 10^{-39} \pi \Delta\varepsilon_{max} w / \nu^0 \qquad (2\text{-}56)$$

Here, w devotes the width of the observed VCD or absorption band. Equations (2-55) and (2-56) hold for Lorenzian band; for Gaussian bands, the factor π needs to be replaced by $\sqrt{\pi}$.

2.3 *Application to the Structural Study of Macromolecules*

Chiroptical techniques are very useful to investigate the solution structure of molecules since they can distinguish molecules that are identical except for their configuration. For macromolecules such as nucleic acid and protein, VCD is more powerful than CD in its ability to distinguish the difference in the secondary structure. In VCD, the effects of the chirality on vibrational, not on electronic transition, are observed, which offers more and narrower spectral features. In addition, the dipole change of these modes occur in a direction that is nearly parallel to the bond connecting the carbon and oxygen atoms. Thus, the infrared VCD observed in the carbonyl stretching vibrations can be interpreted in term of the relative position of the carbonyl groups of the purine and pyrimidine bases of the nucleic acid or the peptide units of protein. With the help of exciton theory, information of the base alignment or peptide orientation can be obtained.

Considers, for example, a dimer which has two C=O groups for each unit (Figure 2-9). According to the DECO mode, we can write down rotational strengths and transition frequencies as functions of the relative position and orientation of these two C=O in the space (Bloomfield, et al., 1974).

$$\nu_{12} = \mu^2 / R^3[\sin\theta_2 \cos(\varphi_1 - \varphi_2) - 3\sin\theta_1 \sin\theta_2 \cos\varphi_1 \cos\varphi_2]$$

$$\nu_{\pm} = \nu_0 + \nu_{12}$$

$$R_{\pm} = \pm(\pi\nu_0 / 2c)R\mu^2 \sin\varphi_1 \cos\theta_2$$

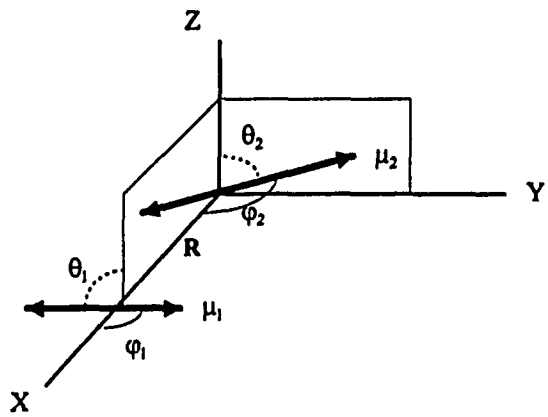


Figure 2-9. The orientation of a dimer in space. The bond joining the two parts of the dimer is along the X-axis and μ_1 is in the XY-plane.

The different orientation of the two C=O groups gives the different VCD pattern. The differences are in frequency shift from the original transition frequency, signal pattern of the two band (positive, negative or negative positive) and magnitude of the intensities.

3. Spectra Measurement and Molecular Structure

3.1 Experimental Considerations

Ultraviolet absorbency measurements were performed by using a Cary Model 219 spectrophotometer. Electronic crcular dchroism measurements were made by using a Jobin-Yvon Mark V Dichrograph. The Mark V CD is equipped with a 250 W xenon lamp and a double prism monochromator. Circularly polarized light is generated through a photoelastic modulator.

All infrared VCD and absorption spectra reported here were collected on a dispersive VCD spectrometer designed and constructed in Dr. Diem's lab. at Hunter college, which has been described in detail [Lee and Diem, 1992; Diem 1993]. The IR absorbance spectrum is taken as a prescan before the VCD scan.

All polyribonucleic acids were obtained commercially (Sigma Chemical). They were lyophilized to dryness from D_2O to remove exchangeable protons, and dissolved to a concentration of between 10 to 40 mg/ml in D_2O / sodium cacodylate buffer solution. Samples are contained between 19 mm diameter CaF_2 plates held apart by Teflon spacers. The exact concentration and path length conditions employed for each of the polymers are given in the Figure captions. All spectra were obtained at 20 °C except those of poly (rG) at low pH, which were collected at 10 °C. The spectra were averages of 30 scans, obtained at a scan speed of 1 cm^{-1} / sec and a 1 sec time constant. Baselines obtained for the sample cells filled with the

solvent/buffer were subtracted from the RNA spectra, to account for a small instrument baseline artifact.

Coordinates for the nucleic acid bases for canonical structures were obtained from the program MacroModel (Still, 1989), operating on an Evans & Sutherland graphics workstation, or were derived from literature data. For small oligonucleotides, canonical geometry are not particularly accurate in describing the nuclear coordinates, but in the polymers reported here, these geometry are adequate.

Our calculations allow us to assume directions of the appropriate transition moments other than the exact bond directions as shown in Figure 3-1. We use AMPAC or other semi-empirical calculations (Tsuboi, et al., 1973) of vibrational spectra to determine the amount of mixing of vibrational coordinates. The AMPAC calculations were carried out on an IBM 6000 RISC station.

VCD and infrared absorption calculations, using the DECO, NECO and PPS approaches, were carried out via FORTRAN programs written in our laboratory, and executing on fast personal computers. The dipole and rotational strength were converted to spectral features by overlaying band shapes as indicated above, and storing the spectra in the same format as the observed data to allow for direct comparison.

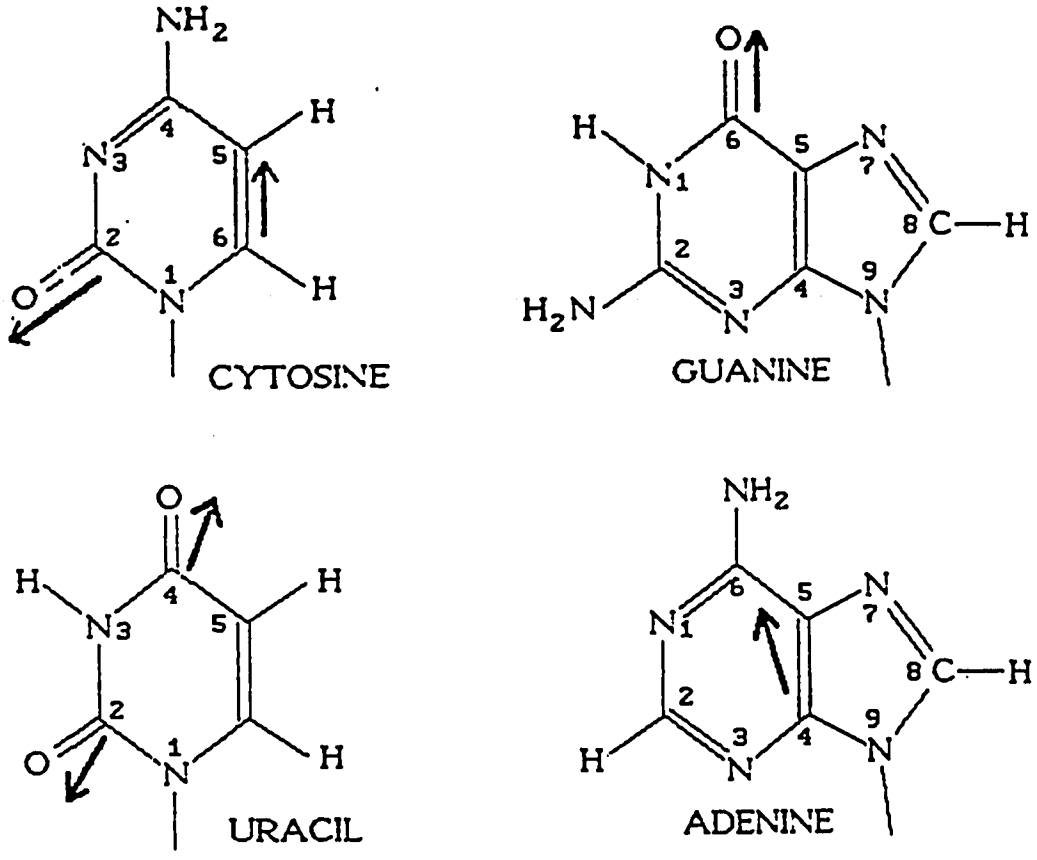


Figure 3-1. Structures of the nucleic acid bases. The directions of the appropriate transition moments are based on AMPAC calculations, and are drawn approximately

Transition dipole moments were derived from the integrated absorption intensities of the corresponding monomers. The monomer vibrational frequencies, intensities and assignments are listed in Table 3-1.

	[a] frequencies (cm^{-1})	[b] ε_{max} (L/mol. cm)	[c] assignment
CMP	1612	450	$C_6 = C_5 - C_4$
	1650	980	$O = C_2$
GMP	1655	980	$O = C_6$
AMP	1623	1100	$C_6 = C_5 - C_4$
UMP	1654	650	$O = C_4 - C_5$
	1687	1100	$O = C_2 - N_1$

Table 3-1. *Vibrational Frequencies of the Ribonucleotide Monomers in D_2O/Sodium Cacodylate Buffer*

In Table 3-1:

a. All frequencies are believed accurate to ±5 cm^{-1}

b. Due to difficulties in determining the accurate cell path length and sample concentration, ε_{max} is taken from Tsuboi et al., 1973

c. The assignment is from Tsuboi, et al., 1973

The magnitude of the calculated spectra need to be scaled by a factor F_{ABS} or F_{VCD} (given in the Figure captions) to match the experimental results. The variance comes from the uncertainties of molar concentrations and pathlengths as well as deviations from the Beer-Lambert's Law. The number of bases used in the calculation also influences the magnitude of calculated VCD and IR spectra.

3.2 Results and Discussion

In the following section, we will present observed VCD and infrared absorption spectra of various single stranded, double stranded polymers, and will compare these observed spectral features with the computed data, using the models introduced earlier.

The nucleic acid is a complex polymer: its unit (base + ribose + phosphate) contains 29-33 atoms. In general, the vibrational spectrum of a polymer with such a big unit is very complex, and a reliable assignment of each vibrational band is not always easy. The calculation of normal vibration often helps our understanding on the nature of infrared absorption bands. In infrared spectroscopy, D_2O is available as a solvent in the 1400 - 1800 cm^{-1} region, and in this spectral region, nucleic acid gives strong absorption bands assignable to in-plane vibrations of the base-residues.

3.2.1 Poly(rC)

We shall start with the experimental and computational results on poly(rC), which exists in buffered aqueous solution in a single stranded right handed helical

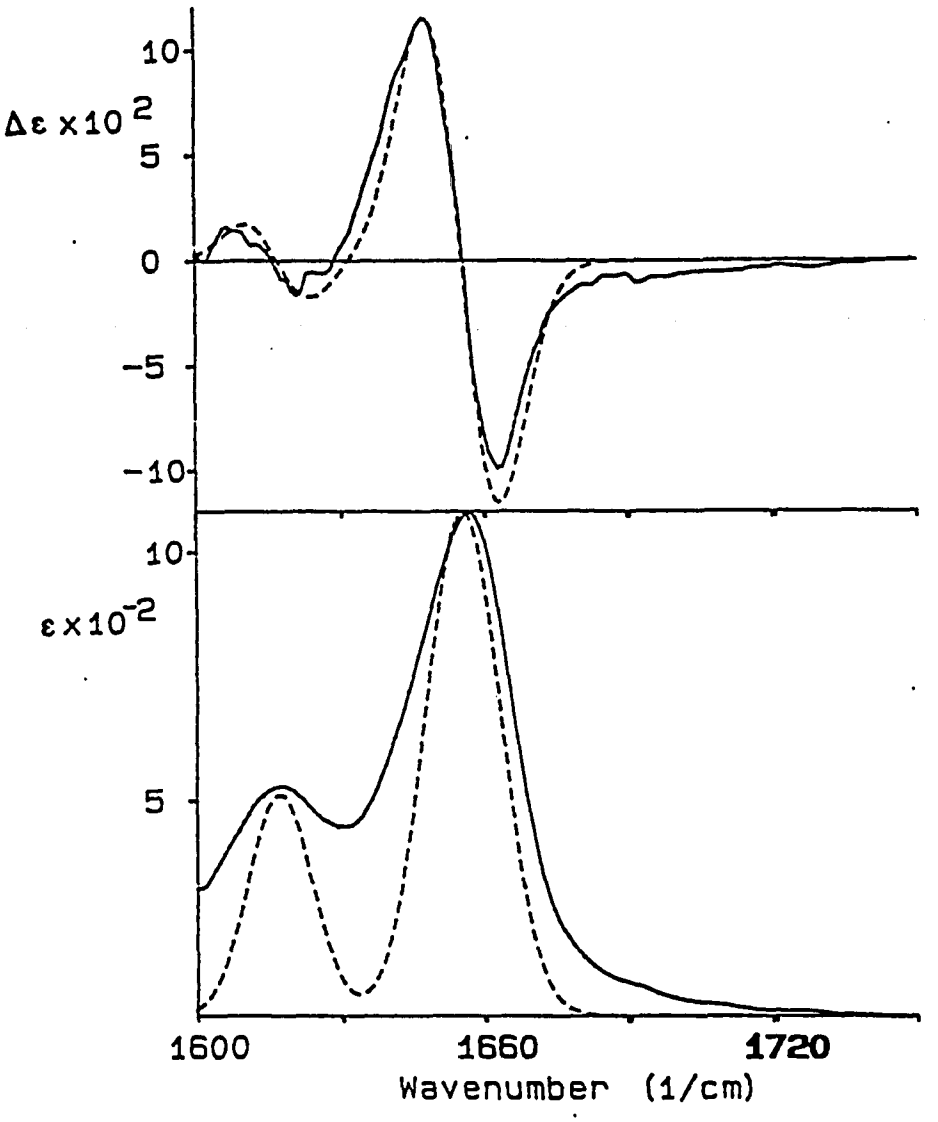


Figure 3-2. Observed (solid traces) and calculated (dashed traces) infrared VCD (top) and absorption (bottom) spectra of poly(rC) in D_2O / sodium cacodylate buffer. Experimental conditions: poly(rC) = 40 mg/ml, T = 20 °C, path length = 50 μm, data expressed in molar extinction units per base. Computation conditions: Two independent DECO computations for 30 bases, result normalized to one base. The coordinates are constructed from Arnott, et. al., 1976. Scale factors are F_{ABS} = 1.13, F_{VCD} = 0.315.

conformation (Arnott, et al., 1976). The experimental VCD and infrared absorption spectra are shown in Figure 3-2. The absorption shows two distinct peaks, at 1617 and 1656 cm^{-1} . Two distinct VCD couplets are observed under the absorption peaks, with the VCD zero crossing occurring at the peak maximum. Both are positive, conservative couplets, which is typical for a right handed helix. Conservative in this context implies that the negative and positive intensities in a couplet are equal; furthermore, we define a positive couplet to be one where the positive lobe is at lower energy (wavenumber) than the negative one.

In order to calculate VCD spectra for a polymer using the calculational models derived in Section 2.2, we need to know the transition frequencies and intensities of the base residues. For the poly(rC), the base residue is cytosine. The infrared absorption of 5'-CMP in a neutral D_2O buffer at 1600-1750 cm^{-1} shows two strong absorption peaks at 1612 cm^{-1} and 1650 cm^{-1}. The transitions have been assigned (Tsuboi, et al., 1973) as follows. The strong peak at 1650 cm^{-1}, which corresponds to the 1656 cm-1 peak in the poly(rC) infrared absorption, is typical for the C=O stretching vibration, whereas the 1612 cm^{-1} peak, which corresponds with the 1617 cm^{-1} peak in the poly(rC), was assigned to a mixture of the $C_5 = C_6$ and the $C_4 - C_5$ stretching coordinates (cf. Figure 3-1). For CMP, the temperature effect on the absorption curves is almost negligible, and no salt effect has been detected. Thus, the two absorption peaks are considered to be caused by the base residues of each nucleotide molecule isolated in D_2O and to be almost free from inter-base

interactions. The CMP infrared spectrum provides us monomer's intensity and transition frequency which we need for the poly(rC) spectra calculation. Beside these, we also need to know the direction of the transition moment, which is related to the change in the molecular dipole moment due to the vibrational motion. The transitions we considered in the 1600 - 1750 cm^{-1} region are all in-plane stretching, so that the direction of the transition moment is determined by the three-dimensional organization of the atoms in each vibrational mode. AMPAC calculations for the 1612 cm^{-1} vibration yielded a potential energy distribution of about 95% $C_5 = C_6$ and 4% C_4 - C_5. Thus, we assumed that the transition moment of this vibration lies along the double bond. Furthermore, we assume that the carbonyl stretching mode (1650 cm^{-1}) lies along the bond connecting the carbon and oxygen atoms. These dipole moments are indicated in Figure 3-1.

Poly(rC) has been shown to exist as a highly ordered asymmetric structure at a neutral buffer solution by CD and ORD (Fasman, et al., 1964). The influence of hydrogen bonding on helical stability is negligible and the forces responsible for maintaining the secondary structure have been suggested to be hydrophobic in nature. It is implied that the helical structure at the neutral pH is single stranded. X-ray fiber diffraction studies have shown that at pH=7.0, poly(rC) can adopt a single-stranded, right-handed, 6-fold helical structure with an axial rise per residue of 0.311 nm. The furanose rings have the C-3'-endo conformation. The other conformational angles of this structure are also remarkably similar to those observed in A-type double stranded

polynucleotide structure. X-ray results showed there is no hydrogen bonding between bases. The molecular structure appears to be stabilized primarily by base-stacking interactions (Arnott, et al., 1976).

The computed VCD spectra are shown as dashed traces in Figure 3-2. Calculations were carried out for a single strand of 30 bases, for which the coordinates were constructed from the X-ray data (Arnott, et al., 1976). The computed spectra are plotted in units of molar coefficients per base to permit a direct comparison with the experimental data.

The spectra were calculated under the assumption that the two transitions at 1617 and 1656 cm^{-1} do not interact. Thus, two separate DECO VCD and infrared absorption spectra were calculated according to equations (2-39), (2-40) and (2-44) for two transitions, and subsequently added. Since both observed VCD spectra are nearly conservative, this assumption appears justified, and the agreement between observed and calculated absorption and VCD band shapes and frequencies confirms the validity of this approach.

Thus, the VCD of poly(rC) in the 1600 to 1750 cm^{-1} spectral region is an example of a system which can be reproduced computationally by considering two non-interacting transitions, and calculating their spectral properties independently.

3.2.2 Poly(rA)

Adenine, shown in Figure 3-1, is the base-residue of poly(rA). Because adenine does not contain a C=O functional group, poly(rA) does not exhibit infrared absorption between 1650 and 1700 cm^{-1} . The peak observed at about 1627 cm^{-1} (Figure 3-3), which corresponds with the 1623 cm^{-1} absorption of adenine, has been assigned to a coupled C_4=C_5 / C_5 -C_6 stretching motion (Tsuboi, et al., 1973). As is the case for cytosine, adenine has no observed salt effect and its 1623 cm^{-1} peak is shifted slightly, with only a small decrease in its intensity on elevating the temperature to 75 °C. Therefore its frequency and intensity can be used for the poly(rA) spectra calculation.

Poly (rA) is able to form double- and single-stranded structures, depending upon the pH of the environment. Under acidic condition, the double-stranded structure is formed. The adenine heterocycles become protonated at N, and two protonated poly(rA) molecules combine *via* base-pairs to yield a double helical complex in which, however, the polynucleotide chains are arranged parallel to each other and not antiparallel. At neutral or alkaline pH, poly(rA) occur as a single-stranded helical structure. Using ultraviolet circular dichroism spectra, the transition from single-stranded to double stranded can be observed by measure the spectra at different pH conditions (Oikawa, et al., 1969). Our VCD and IR spectra calculation is based on a single stranded structure with an adjusted transition moment direction shown in Figure 3-1, which is 73% C_4=C_5 + 27% C_5-C_6. The structural features of

poly(rA) were derived from MacroModel program, using single stranded RNA parameters. VCD calculations were carried out for a 30 base single strand, using a ε_{max} value given by Tsuboi, et al., (1973). The agreement between the calculated and experimental spectra confirms the single stranded model for poly(rA) at the experimental conditions.

The VCD spectrum of this vibration is conservative as well, and the VCD zero crossing occurs at the absorption maximum. Computational results based on the DECO approach reproduce this behavior well, and are shown in Figure 3-3. Thus, the calculated and observed results on poly(rA) and poly(rC) indicate that a simple exciton type mechanism is responsible for the vibrational optical activity of these polymers in both C=O and C=C stretching vibrations. In both cases there is good qualitative agreement between theoretical and experimental results, in both the C=O stretching modes as well as the less polar C=C-C stretching mode.

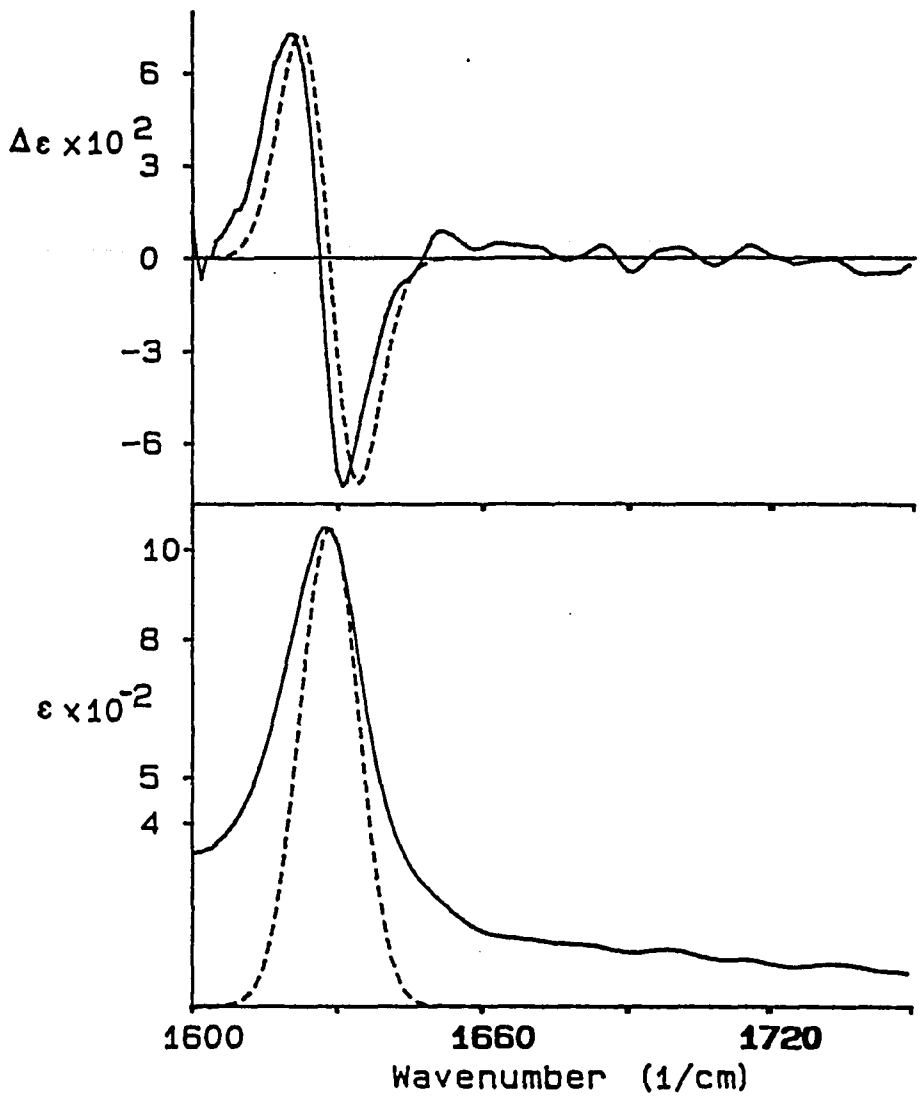


Figure 3-3 Observed (solid traces) and calculated (dashed traces) infrared VCD (top) and absorption (bottom) spectra of poly(rA) in D_2O / sodium cacodylate buffer. Experimental conditions: poly(rA) = 30 mg/ml, T = 20 °C, path length = 50 μm, data expressed in molar extinction units per base Computation conditions: DECO computations for 30 bases, result normalized to one base. Scale factors are F_{ABS} = 0.963, F_{VCD} = 0.385.

3.2.3 Poly(rG)

Guanine residues have been observed to be particularly prone to self-aggregation, even in short oligonucleotides. In a immunological study of poly(rG), the evidence has been found supporting the simultaneous existence of two forms of this polymer in solution; a single- and a four-stranded form (Souleil, et al., 1968). CD calculations have been done on single- and four-stranded poly(rG), and the comparison of calculated poly(rG) CD spectra for a single-stranded A-RNA model and a four-stranded model with the measured spectrum of poly(rG) has been presented (Williams, et al., 1983). The measured poly(rG) spectrum has a long-wavelength positive shoulder from 305 to 270 nm, followed by an intense positive band around 260 nm with a crossover point at 250 nm and a negative peak at 235 nm. The calculated single-stranded A-RNA spectrum is negative at long wavelengths, crossing over at 267 nm to a intense positive band at 254 nm. The calculated four-stranded poly(rG) spectrum has a positive long-wavelength band at 280 nm, a positive minimum at 265 nm, and a positive maximum at 245 nm. The positive band found at long-wavelengths in the measured poly(rG) CD spectrum could arise from the presence of a greater amount of four-stranded structure in solution than single-stranded. Williams (Williams, et. al. 1983) assumed the ratio of single- to four-stranded poly(rG) to be 40% : 60% in the CD calculation to get the agreement with the measured poly (rG) spectrum.

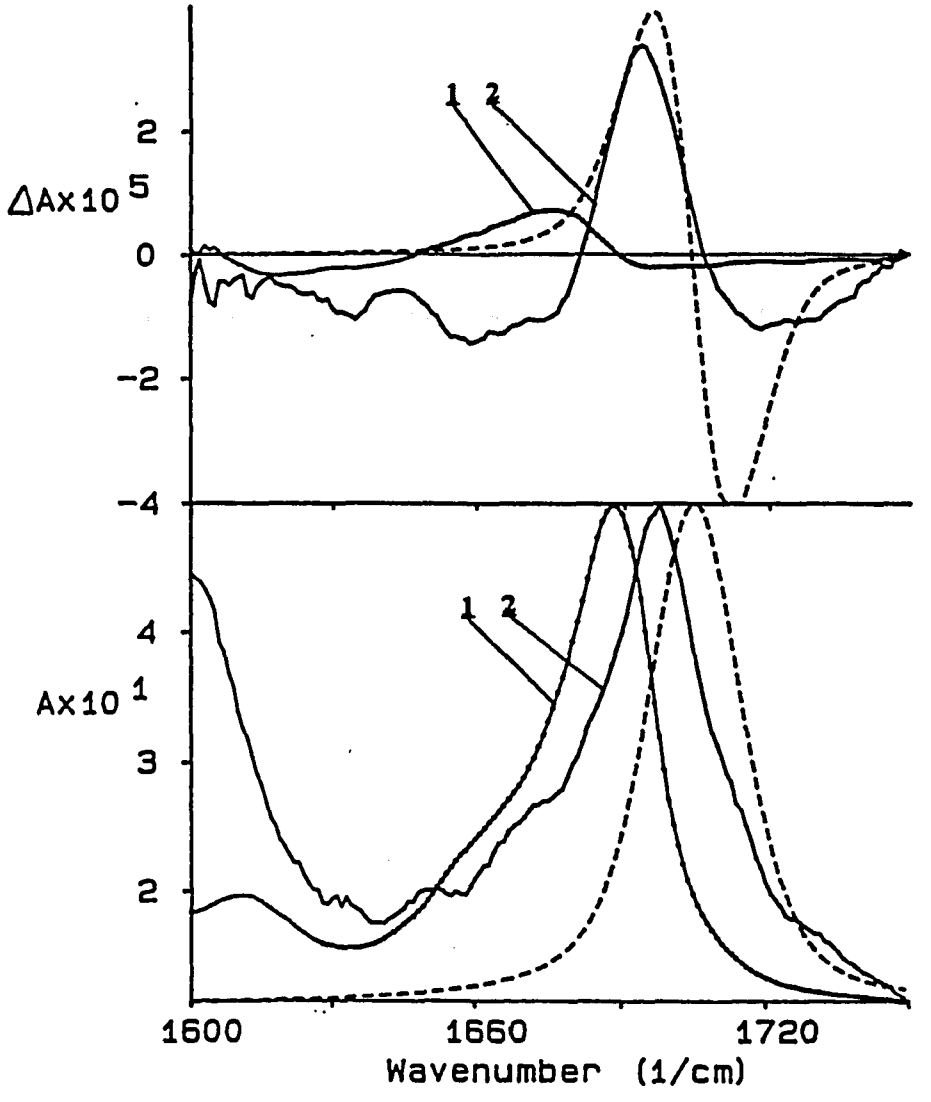


Figure 3-4. Observed and calculated infrared VCD (top) and absorption (bottom) spectra of poly(rG) in D_2O / sodium cacodylate buffer. Solid line 1: pH=7, T=10 °C. Solid line 2: pH=3 and sample was heated to 70 °C for 40 minutes then quickly quenched to -150 °C for 1 hour. The spectrum was taken at 10 °C. Dashed line: single stranded DECO calculated result. We use A and ΔA instead of ε and Δε for the difficulty to determine the concentrations of the samples. Comparison was based on the IR intensity which were kept the same. Scale factor $F_{VCD} = 0.1$.

Our VCD spectra also indicated this differently stranded poly(rG) at different experimental conditions. From Figure 3-4 one can see that at neutral pH, the observed and calculated spectra (using a single stranded geometry) of poly(rG) do not agree at all. However, it is well established that poly(rG) exists at room temperature and pH of 7 in multi-stranded form. When the pH is lowered to 3 and the temperature is raised to 70 °C or higher, poly(rG) begins to melt towards a single stranded conformation.

The UV CD in Figure 3-5 to Figure 3-7 indicated this transition. At pH=7 and room temperature (Figure 3-5), the CD spectrum shows multi-stranded poly(rG) which has a positive peak at 290 nm. At pH=3 and a temperature of 70 °C, the partial melting of multi-stranded poly(rG) to a single stranded form occurs, which is indicated by a negative couplet at 285 nm, a crossing over at 275 nm and the intense positive peak shift towards low-wavelength (Figure 3-6). With a temperature increase to 90 °C, the magnitude of the negative 285 nm peak increases (Figure 3-7). At this point the spectra are almost the same as the single stranded spectrum described by the CD calculated result (Williams, et al., 1983).

In order to measure the VCD spectra of poly(rG) in its single stranded form, ideally, one would like to maintain the temperature above 90 °C for some time to insure a complete phase transition. However, the infrared cell used will not allow a temperature this high to be reached without significant sample leakage; thus, the

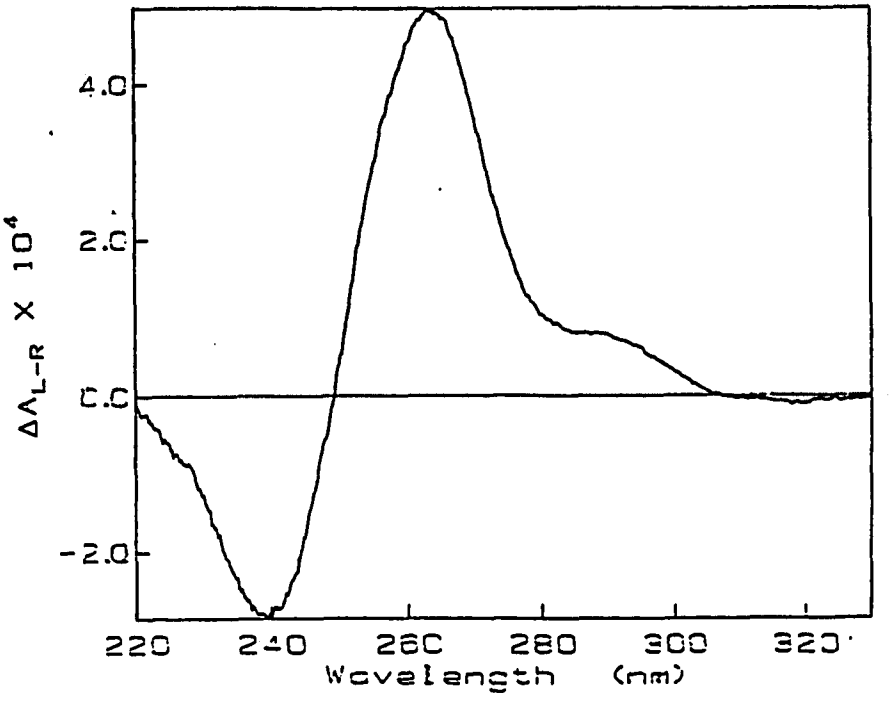


Figure 3-5. CD spectrum of poly(rG) at 25 °C and pH = 7.

samples were heated to over 70 °C for 40 minutes and then quickly quenched to -150 °C and kept at -150 °C for 1 hour. The VCD and IR spectra were taken at 10 °C.

The multistranded poly(rG) shows an infrared absorption at 1688 cm^{-1}, with a low frequency shoulder at about 1660 cm^{-1}. The VCD signal of this form is small, and consists of a broad negative intensity at 1615 cm^{-1}, positive VCD at 1675 cm^{-1}, and slightly negative VCD at 1700 cm^{-1} (cf. Figure 3-4-#1). The high frequency zero crossing of the VCD trace occurs at the absorption maximum. Using published

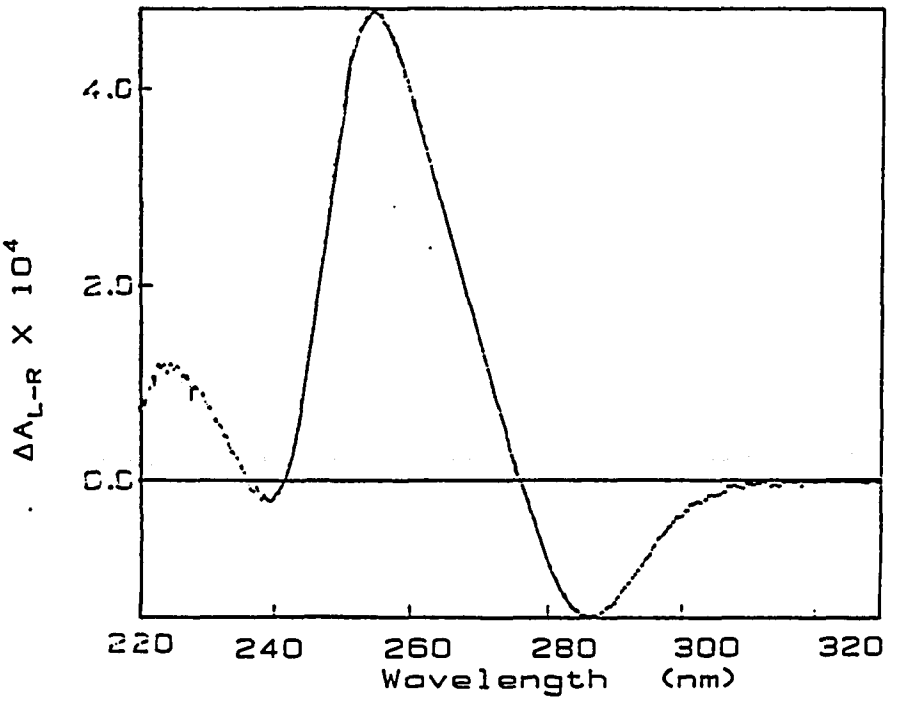


Figure 3-6. CD spectra of poly(rG) at 70 °C and pH = 3.

coordinates (Arnott, 1974) for the quadruple strand, the main VCD couplet was predicted with the wrong sign pattern.

After heating poly(rG) at low pH to above 70 °C and cooling to -150 °C, the spectra show in Figure 3-5-#2 are obtained. Compared with CD spectra in Figure 3-7, at this condition, the polymer is mostly single stranded, with a small fraction of multiple-strands. The positive VCD band of the polymer is narrower and more intense. The infrared absorption band of the polymer is shifted to 1698 cm^{-1}. The VCD and IR spectra of this partial single stranded poly(rG) can be reproduced

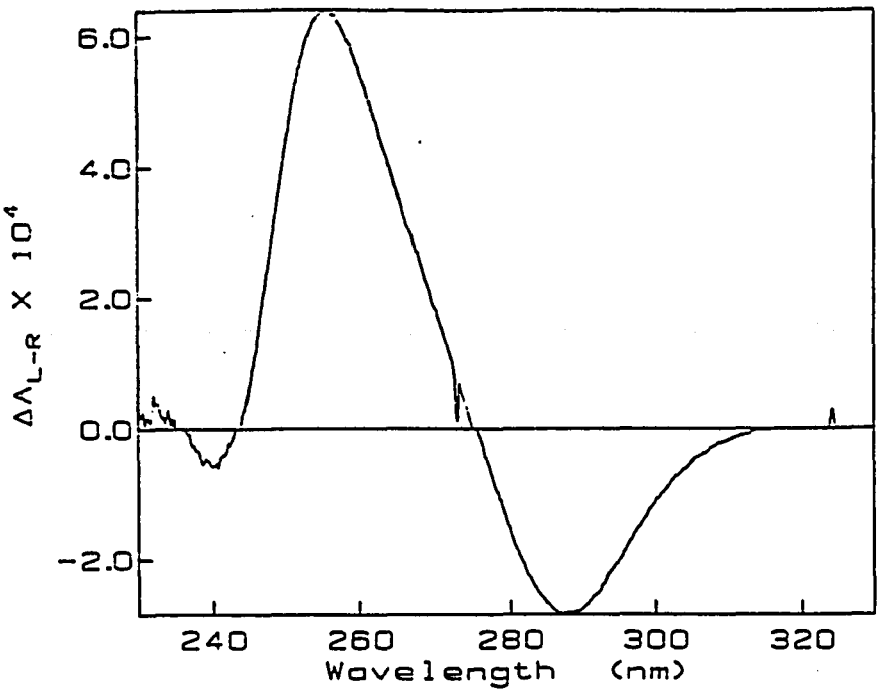


Figure 3-7. CD spectra of poly (rG) at 90 °C and pH = 3. Because the instrument is difficult to run at T > 70 °C, the sample was heated to 90 °C for 40 minutes and quickly quenched to -150 °C for 1 hour. the spectrum was taken at 25 °C.

reasonably well with a single stranded DECO calculation.

3.2.4 *Poly(rG)·Poly(rC)*

Poly(rG)·poly(rC) exists in buffered, low ionic strengths (0.05M NaCl) aqueous solution in a double stranded, right handed conformation belonging to an A-family of conformations. The observed and calculated VCD and absorption spectra are shown in Figure 3-8. The observed spectra are somewhat different from those observed for B-form poly(dG)·poly(dC) [Zhong et al., 1990]; thus, it appears that A and B form polymers can be distinguished *via* VCD (Wang & Keiderling, 1992). The main features of poly(rG)·poly(rC) are three absorptions at 1621, 1649 and 1690 cm^{-1}, whereas in poly(dG)·poly(dC), the two high frequency peaks overlap, and the low frequency peak is observed only as a shoulder. The VCD spectrum of the ribonucleic acid consists of three couplets: a weak positive one with a zero crossing at 1617 cm^{-1}, a weak negative one with a zero crossing at 1636 cm^{-1}, and the main positive-negative signal with a zero crossing at 1691 cm^{-1}. This signal is positively biased. The corresponding VCD signal of poly(dG)·poly(dC) is much more conservative.

It is interesting to note that guanine and cytosine monomers have their absorption peaks below 1660 cm^{-1}, at 1655 and 1650 cm^{-1}, respectively. In poly(rG)·poly(rC), however, the highest wavenumber peak is observed at 1690 cm^{-1}. Thus, this peak must be due to interaction of the C=O dipoles, which shifts some exciton components to higher and some others to lower frequencies. This behavior is reproduced well by the calculated results, depicted in Figure 3-8. It is also noteworthy that for guanine and adenine, although they have relatively similar structures, the

vibrations in the 1600 to 1650 cm^{-1} region are quite different. Adenine, devoid of a C=O group, has a strong C=C-C stretching vibration at 1623 cm^{-1}, with a dipole moment comparable of that of a C=O stretching vibration. Guanine, on the other hand, only has the C=O stretching vibration in the 1600 to 1700 cm^{-1} region, and the corresponding C=C-C vibration occurs below 1600 cm^{-1} (Tsuboi, et al., 1973).

The calculated VCD spectra were obtained using C=O stretching for guanine (1655 cm^{-1}) and cytosine (1650 cm^{-1}) and C=C-C stretching mode on cytosine (1612 cm^{-1}). Figure 3-8 is the comparison of the observed and the calculated VCD and IR spectra of poly(rG)·poly(rC). From Figure 3-8 one can see that the DECO calculation (#1 dashed traces) reproduced the VCD intensity not quite as well, and predicted the main infrared absorption and zero crossing of the main couplet at significantly lower wavenumber. PSS calculations were also carried out, using the three vibrations enumerated above as the transitions in one subunit. The PSS calculations (#3 dashed traces) reproduced the low frequency VCD slightly better than the NECO calculations, but reproduced the infrared absorption intensities less accurately. The NECO calculation (#2 dashed traces) gives the best result.

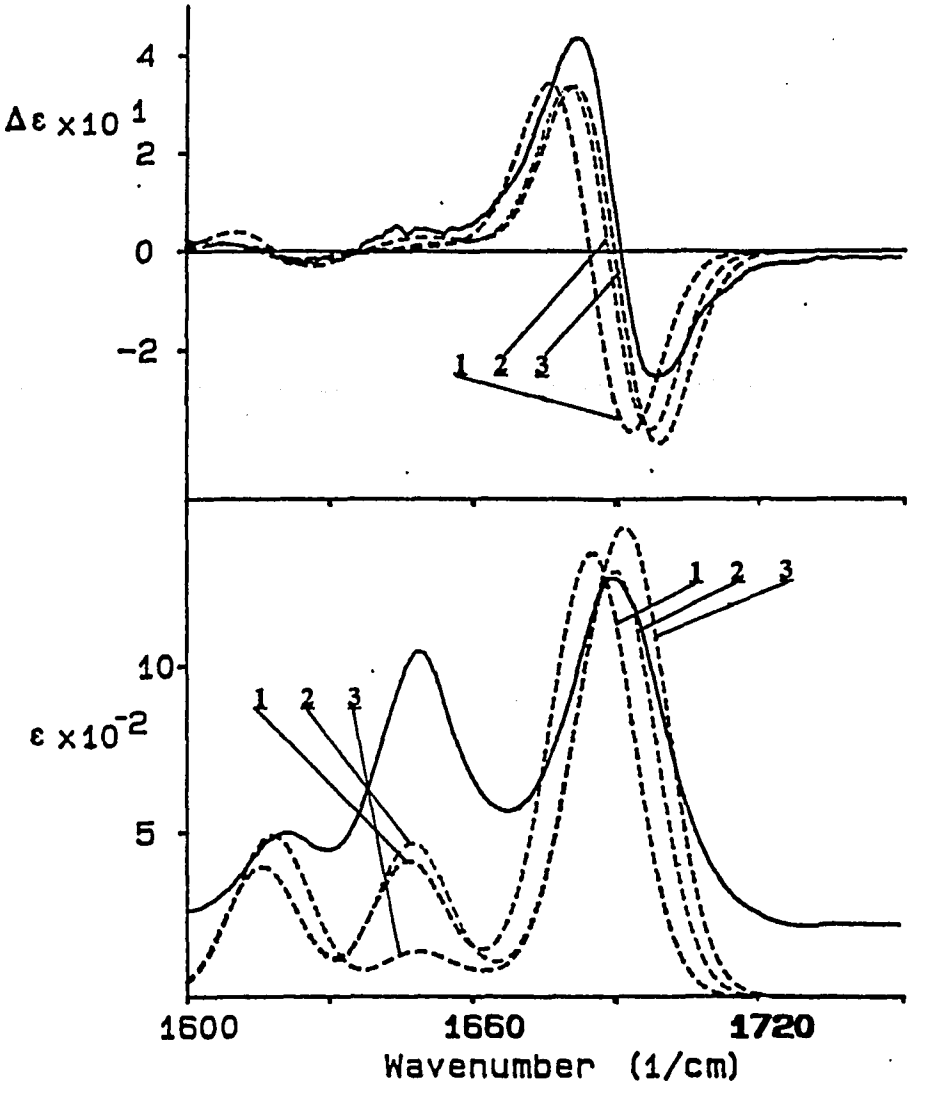


Figure 3-8. Observed (solid traces) and calculated (dashed traces) infrared VCD (top) and absorption (bottom) spectra of poly(rG).poly(rC) in D_2O / sodium cacodylate buffer. Experimental conditions: poly(rG).poly(rC) = 35 mg/ml, T = 20 °C, path length = 50 μm, data expressed in molar extinction units per base. 1: DECO computation for C=O groups, independent DECO calculation for C=C-C vibration. 2: NECO computation for C=O groups, independent DECO calculation for C=C-C vibration. For 1 and 2, 15 base pairs, results normalized to one base pair, F_{VCD} = 0.470. 3: PSS computation, 10 base pairs, results normalized to one base pair, F_{VCD} = 0.500.

3.2.5 *Poly(rA)·Poly(rU)*

X-ray diffraction patterns show the synthetic complementary RNA double-stranded poly(rA).poly(rU) exists in A-RNA right-handed double helix with 11-base - pairs per turn.. On increasing the salt content of these synthetic RNA fibers, a transition to a similar right-handed, 12-base-pair per turn helix (A'-RNA) is observed (Arnott, et al., 1972). Our calculation is based on an A-form RNA geometry.

Calculations on uracil-containing polymers are more difficult to carry out, since the two C=O groups on the U residue are in such close proximity that their coupling is exaggerated in standard DECO calculations. The DECO model has been successfully used in the DNA polymer VCD and IR spectra computation (Zhong, et al., 1990), but the attempt to calculate poly(dA-dT)·poly(dA-dT) spectra failed because the poly(dA)·poly(dT) spectra suffered from a mis-assignment of the two C=O frequencies. Here, we wish to report a corrected calculation for poly(rA)·poly(rU). The only vibration of the A residues is taken to be identical to the one described above for poly(rA). However, the two carbonyl stretching vibrations in U require special attention, since they appear to mix more with base stretching modes than C=O modes on other nucleotides. The AMPAC calculations give the following contributions of ring vibrations to the modes observed at the wavenumbers listed (cf. Figure 3-1):

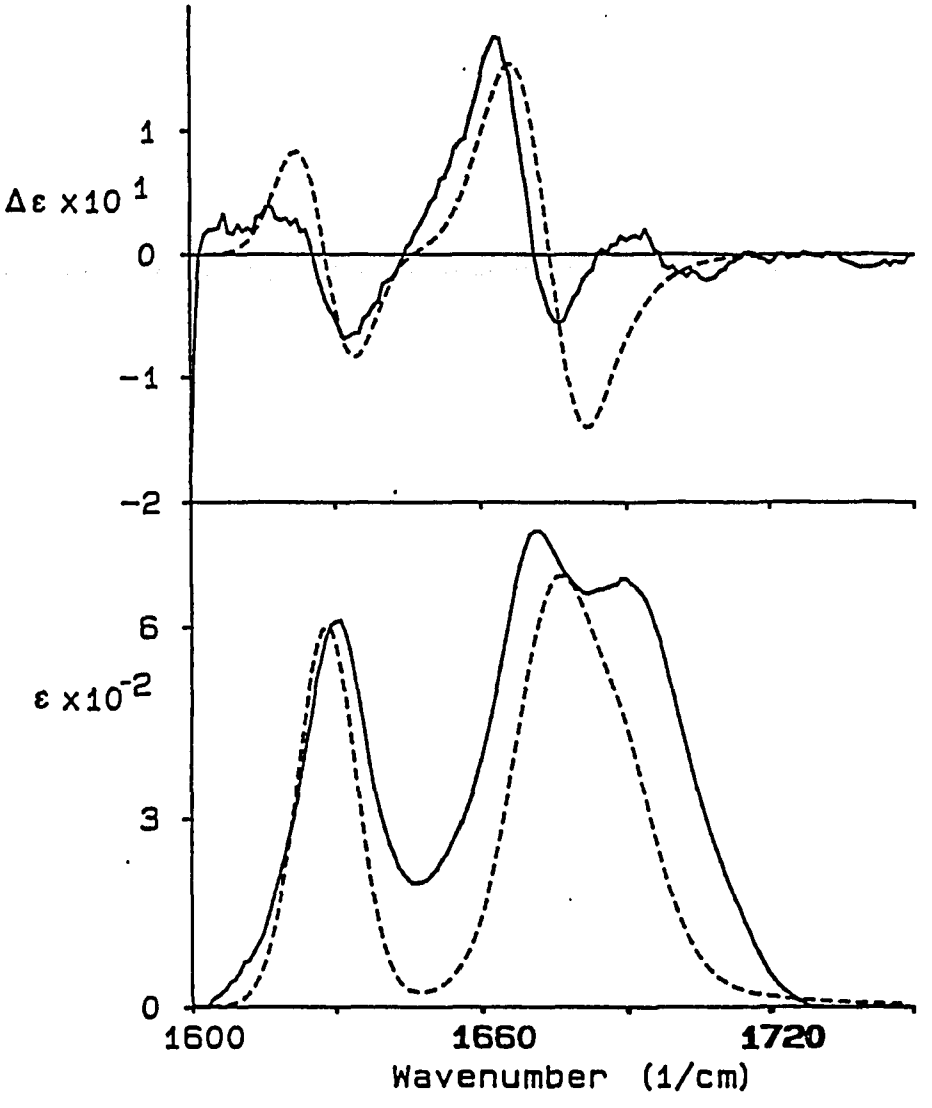


Figure 3-9. Observed (solid traces) and calculated (dashed traces) infrared VCD (top) and absorption (bottom) spectra of poly(rA).poly(rU) in D_2O / sodium cacodylate buffer. Experimental conditions: poly(rA).poly(rU) = 30 mg/ml, T = 20 °C, path length = 50 μm, data expressed in molar extinction units per base. Three independent DECO calculations for C=C-C, O=C-C and O=C-N vibrations, 20 base pairs, results normalized to one base pair, F_{VCD} = 0.550, F_{ABS} = 0.470.

$O=C_2-N_1$: 1687 cm^{-1} 86% $O=C_2$, 14% C_2-N_1 (uracil)

$O=C_4-C_5$: 1654 cm^{-1} 79% $O=C_4$, 21% C_4-C_5 (uracil)

$C_4=C_5-C_6$: 1625 cm^{-1} 73% $C_4=C_5$, 27% C_5-C_6 (adenine)

The approximate directions of the dipole moments are given in Figure 3-1. Since these three vibrations are relatively far apart, interactions between them were expected to be relatively small, and therefore, independent DECO calculations were carried out for all three vibrations. The results for poly(rA)·poly(rU) are shown in Figure 3-9, and the agreement between observed and calculated spectra is very good.

3.3 Summary

In the first part of this thesis, we present the models of VCD and IR intensities which are based on the exciton approach. The models include the degenerate extended coupled oscillator model (DECO), nondegenerate extended oscillator model (NECO) and pseudo-single stand approach (PSS). The effect of including near degenerate vibrations and geometric considerations is discussed. We also show the necessity to include vibrations other than the C=O stretching motion. Our computation was improved by adjusting, using semi-empirical quantum

mechanical normal coordinate results, the direction of the dipole derivative with respect to the C=O bond direction. The observed and computed IR and VCD spectra of a number of polyribonucleic acids in aqueous solution in 1600-1750 cm^{-1} spectra region were reported. The experimental data are compared with results calculated using different levels of sophistication within the exciton approach. Single stranded, right-handed helical poly(rC) shows two distinct IR peaks and two positive, conservative VCD couplets with VCD zero crossing occurring at the peak maximum. The spectra are calculated using the DECO model under the assumption that the two transitions (C=C-C and C=O) do not interact. Poly(rA), which does not have a C=O function group, exhibits a single absorption in its IR spectrum and a single conservative VCD pattern. The peak is due to the C=C-C stretching motion. Its spectra are reproduced well by the DECO model and single stranded RNA parameters with an adjusted transition moment direction. Poly(rG) has two conformations, single-stranded and quadruple-stranded, in solution depending on the pH and temperature. The phase transition can be seen with CD and VCD spectra at different experimental conditions. Poly(rG).(poly(rC) exists as a double stranded right-handed form belonging to an A-comformation. The observed VCD is different from that of B-form poly(dG).(dC). The coupling among the nondegenerate transitions leads to an IR absorption and VCD zero crossing shift. The NECO computation gives the best simulation of the observed spectra. In the calculation of poly(rA).(rU), we adjust the directions of vibrational modes in the U base according to the AMPAC semi-

empirical calculation result and three independent DECO calculation are carried out. The agreement between the observed and calculated spectra is generally good.

Part II
Spectroscopic Analysis of the Tat-TAR interaction

4. Biological and Structural Properties

In 1981, the first cases of a new disease called acquired immune deficiency syndrome (AIDS) were recognized. The cause was identified two years later by Luc Montagnier and Robert Gallo (Stryer, 1988). AIDS is produced by human immunodeficiency virus (HIV), a retrovirus.

Retroviruses are RNA-containing viruses. They encode their genetic information as RNA and thus must undergo reverse transcription by an enzyme called reverse transcriptase into a DNA copy, which is then integrated into the genome of the host.

The HIV-1 long terminal repeat (LTR) contains promoter elements responsible for initiation of viral transcription. After synthesis, the full-length viral RNA is either transported directly into the cytoplasm and becomes translated into structural proteins, or it is spliced into subgenomic RNAs that encode the several viral regulatory proteins (Felberg, et al., 1989; Hammarskjold, et al., 1989). These early transcripts encode two short regulatory proteins, Tat and Rev. Tat is localized within the nucleus and specifically increases the levels of viral mRNAs and protein in a positive feedback loop; Rev protein, specifically stimulates the accumulation of

unspliced and single spliced viral mRNAs. Thus, Rev negatively regulates the synthesis of both itself and Tat (Sharp, et. Al., 1989).

HIV-1 Tat (Trans-activator) protein (Figure 4-1) is absolutely essential for HIV transcription and consequently for viral replication. The increase in viral mRNA and protein synthesis stimulated by Tat is greater than 100-fold, which can be roughly divided into a 20-fold increase in viral mRNA synthesis and a 5-fold increase in the amount of protein synthesized per mRNA. Although these relative values vary significantly owing to different protocols and cell lines, the overall impression is that Tat increases both the level of viral mRNA and its efficiency of translation.

```
                                        10                                  20
Met Glu Pro Val Asp Pro Arg Leu Glu Pro Trp Lys His Pro Gly Ser Gln Pro Lys Thr
                                        30                                  40
Ala Cys Thr Asn Cys Tyr Cys Lys Lys Cys Cys Phe His Cys Gln Val Cys Phs Ile Thr
                                        50                                  60
Lys Ala Leu Gly Ile Ser Tyi Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro Pro Gln
                                        70                                  80
Gly Ser Gln Thr His Gln Val Ser Leu Ser Lys Gln  Pro Thr Ser Gln Ser  Arg Gly  Asp
Pro Thr Gly Pro Lys Glu
```

Figure 4-1 HIV-1 Tat protein. Basic domain is in bold

The LTR region needed for Tat-mediated transregulation has been termed the TAR (Trans-activation response element). Extensive mutagenesis (Hauber, et. al., 1987) has defined a minimal 25 nucleotides, ΔTAR, Figure 4-2.

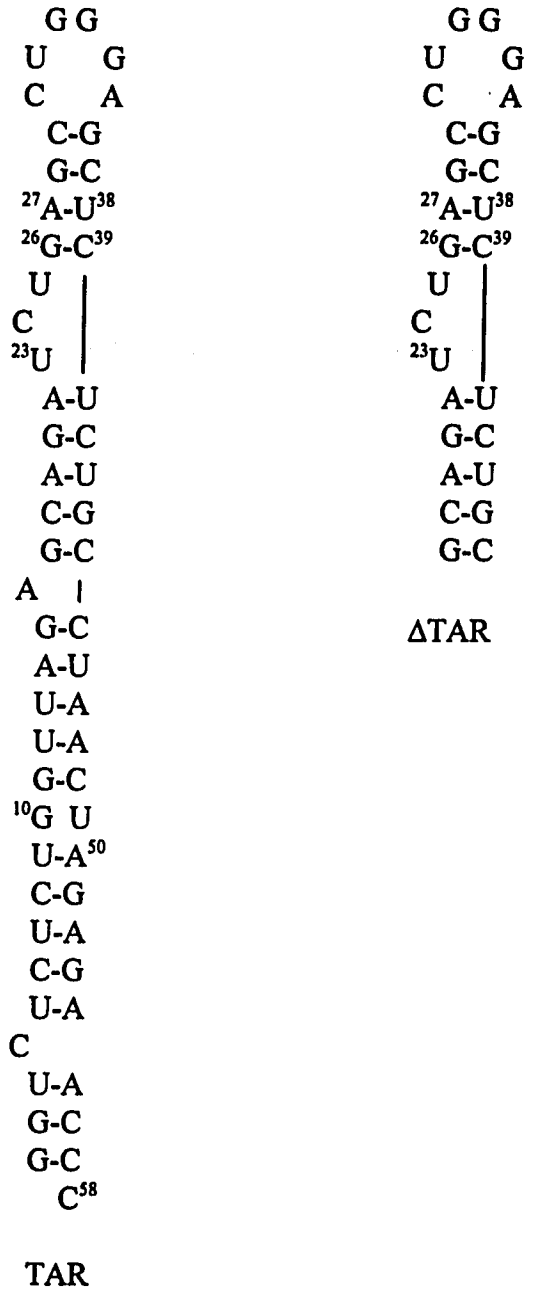


Figure 4-2 TAR and ΔTAR

Tat is 86 amino acids long and contains a highly conserved cysteine-rich region (with 7 cysteines in 16 residues) and a highly conserved basic region (with 2 lysines and 6 arginines in 9 residues). Studies with peptides have shown that TAR RNA binding domain of Tat is this basic region (Weeks, et al., 1990 ; Cordingley, 1990; Calnan et al., 1991b), which appears to recognize and bind TAR with specificity similar to the intact protein, thus providing a relatively simple system for the investigation of TAR-Tat interaction.

A considerable amount is known about TAR RNA structure and the requirements for Tat binding. Nuclease mapping experiments first showed that TAR forms a stable stem-loop structure (Muesing et al., 1987). Sequent deletion and analysis showed that the upper part of the stem-loop (nucleotide 19 to 42 relative to the start of HIV transcription) was sufficient for the Tat response in vitro (Jakobovits et al., 1988). The TAR hairpin contains a six-nucleotide loop and a three-nucleotide bulge that are both essential for Tat activity. Tat binding appears to be limited to the region around the bulge (Dingwell, et al., 1990; Cordingley, 1990; Weeks et al., 1990). Although the size of the bulge is not critical for Tat binding, there is an important requirement for uridine at the 23 position of the bulge (Weeks et al., 1991). The identity of several base pairs surrounding the bulge is also important; in particular, the two base immediately above the bulge G^{26}-C^{39} and A^{27}-U^{38}, are essential (Weeks, et al., 1991).

Using gel electrophoretic methods, the relative binding affinities of ΔTAR with Tat-peptide (14 to 38 residue peptide derived from Tat that spans the basic region) have been measured. K_d is at the level of 16 - 80 pM (Weeks, et al., 1991; 1992; Long & Crothers, 1995). At high RNA concentrations, virtually all of the RNA was bound at 1:1 peptide/RNA stoichiometry, which suggested that one peptide binds per TAR molecule (Calnan et al., 1991). Further studies have shown that the free amino acid arginine binds specifically to the same site in TAR as does the Tat peptide (Tao & Frankel, 1992). Using NMR, the conformation of TAR and the arginine-TAR complex were determined. In unbound TAR, there are three 'bulge' nucleotides that are partially stacked between two A-form RNA stems. In arginine-TAR complex, the bulge nucleotides become unstacked as the two helical stems coaxially stack and a critical uridine (U^{23}) in the bulge is re-positioned in the major groove of the upper stem, potentially forming a base triple with A^{27}-U^{38} base pair, which stabilizes the arginine hydrogen bonding to G^{26} and phosphate (Puglisi, et al., 1992).

In the following section, we will present the CD and VCD measurements of the Tat-peptide, the ΔTAR, their 1:1 complex and ΔTAR-arginine 1:1 complex. We will also show the prediction of Tat-peptide structure by VCD calculation and the kinetic as well as molecular modeling studies of this system.

5. Materials and Methods

5.1 Peptide and RNA Synthesis

Tat-peptide with a tryptophan at the NH_2-terminus and an alanine at COOH-terminus was synthesized using solid phase chemistry and was purified by preparative reverse-phase HPLC. The concentration was determined by tryptophan absorbance at 278 nm.

ΔTAR RNA was prepared by cell-free transcription using T7RNA polymerase and synthetic oligonucleotide templates (Milligan, et al., 1987).

5.2 CD and VCD Spectroscopy

The principle and instrumentation for CD and VCD spectroscopy have been illustrated in Section 2 of Part I. Here, we will introduce the experimental condition and resultant spectra.

All CD spectra reported here were collected on the JOBIN YVON Mark V spectrometer. The Infrared VCD and absorption spectra were collected on the first of two VCD spectrometers at Hunter College. All samples were lyophilized to dryness from D_2O to remove exchangeable protons, and dissolved to the desired concentration in D_2O/ Tris buffer (10 mM Tris-HCl, pH = 7.5,·70 mM NaCl, 0.2 mM

EDTA, 5% glycerol. Calnan, et al., 1991). Samples were contained between 19 mm diameter CaF_2 plates held apart by 25 μm Teflon spacers. All the VCD spectra were collected at 10 °C and CD spectra were taken at room temperature.

5.3 Stopped-Flow Fluorescence Kinetics

Luminescence is the emission of photons from electronically excited states. Luminescence is divided into two types, depending upon the nature of the ground and the excited states. Fluorescence is the emission which results from the return to the lower orbital of the paired electron. Phosphorescence is the emission which results from transition between states of different multiplicity, generally a triplet excited state returning to a singlet ground state. Figure 5-1 is the Jablonski Diagram which illlustrates the absorption and emission of light. The emission from the thermally equilibrated excited state S_1 to the ground state is called fluorescence.

Tryptophan is the most highly fluorescent amino acid in proteins. The tryptophan residues of proteins generally account for about 90% of the total fluorescence from proteins. This natural fluorophore is highly sensitive to the polarity of its surrounding environment. Frequently, spectral shifts and quenching are observed as a result of several phenomena, such as binding of ligands, protein-protein association, and denaturation, among others. In addition, the emission

maximum of protein reflects the average exposure of their typtophan residues to the aqueous phase. Tryptophan absorbs light near 280 nm, and fluorescence emission maximum is around 350 nm. The fluorescence life time of tryptophan residues ranges from 1 to 6 nsec.

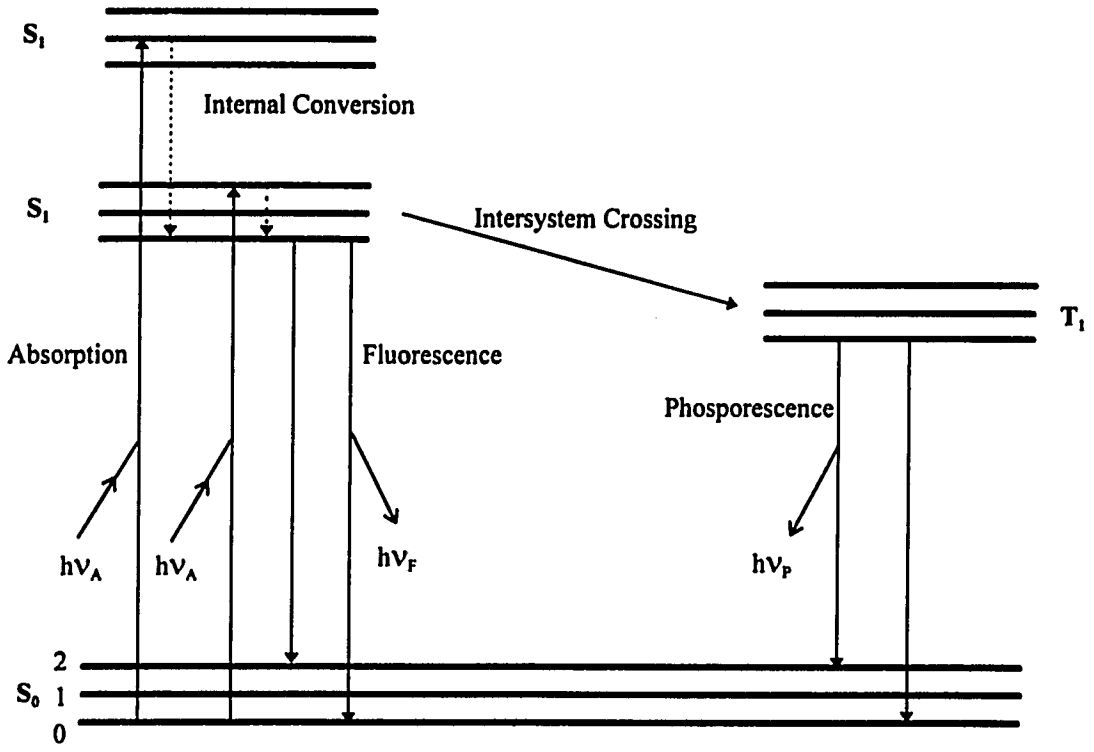


Figure 5-1 Jablonski diagram

The kinetics of Tat-TAR interaction were monitored on a SPEX Fluorolog - τ_2 Spectrofluorimeter and Hi-Tech Scientific SFA-11 Stopped Flow Apparatus. The

spectrofluorimeter consists of five major components: (1) lamp, a 450 W ozone-free Xenon arc lamp; (2) excitation monochromator, a modified Czerny-Turner single-grating monochromator; (3) modulator compartment; (4) sample compartment; (5) emission monochromator. The fluorescence intensity as a function of time was collected and processed by a Gateway 2000 computer (SPEX Fluorolog-τ2 Operation Manual 4-7). Figure 5-2 is the schematic diagram of the stopped flow accessory, which is used to measure the rate constant for a reaction in solution. This involving the mixing of reagent solutions followed by spectrophotometic monitoring of the reacting mixture. The reagent solutions, stored in a pair of 10-mL reservoirs, are admitted to the 2-mL drive syringes through 3-port valves. The valves are turned to connect the drive syringes to the flow circuit and 0.2 mL of solution is rapidly expelled from each syringe by a manually pushing on the drive bar. These incoming solutions displace thermostatted solutions in two inlet tubes which are surrounded by a flexible outer tube carrying the thermostatted fluid. The thermostatted solutions mix in the cell displacing the contents left from a previous run. The solution expelled from the cell passes through a third tube back to the drive unit and fills a 2-mL stopping syringe. When this syringe piston hits a mechanical stop, the flow stops. The observation cell is filled with fresh-mixed thermostatted solution and the consequent reaction is monitored by the spectrofluorimeter. Samples, which were in pH = 7.5 Tris-buffer, were excited at 292 nm and the emission monitored at 350 nm. The data were collected at 10 °C. The rate constants for the reaction of binding of ΔTAR with

Tat-peptide were measured by performing the experiment using at least 5-fold higher concentration of ΔTAR than the concentration of Tat-peptide to maintain the

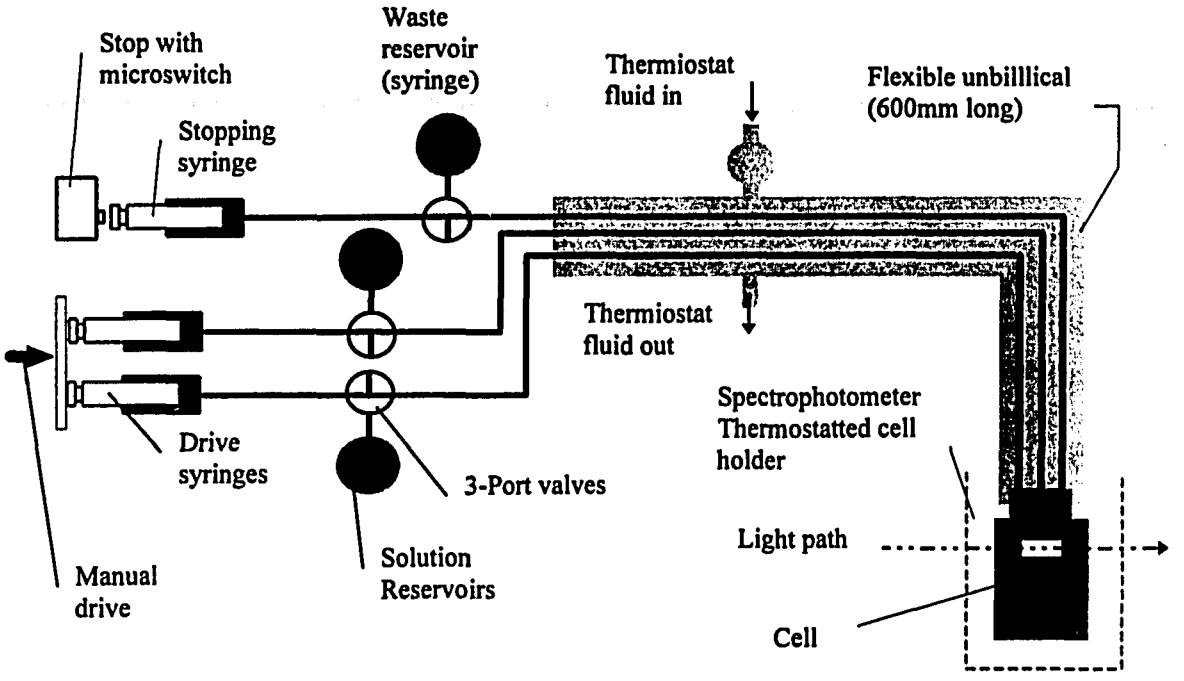


Figure 5-2. Schemaic diagram of the Hi-Tech Scientific SFA-11 rapid kinetics assessory.

pseudo-first order kinetic conditions. The fluorescence of the peptide tryptophan decreases upon binding to the ΔTAR RNA.

6. Result and Discussion

6.1 CD Spectra

Our CD results are generally consistent with the CD results of the Frankel group (Tan & Frankel, 1992). The UV-CD spectra of double-stranded RNA are highly sensitive to the conformation of the polynucleotide helix. In Figure 6-1, the

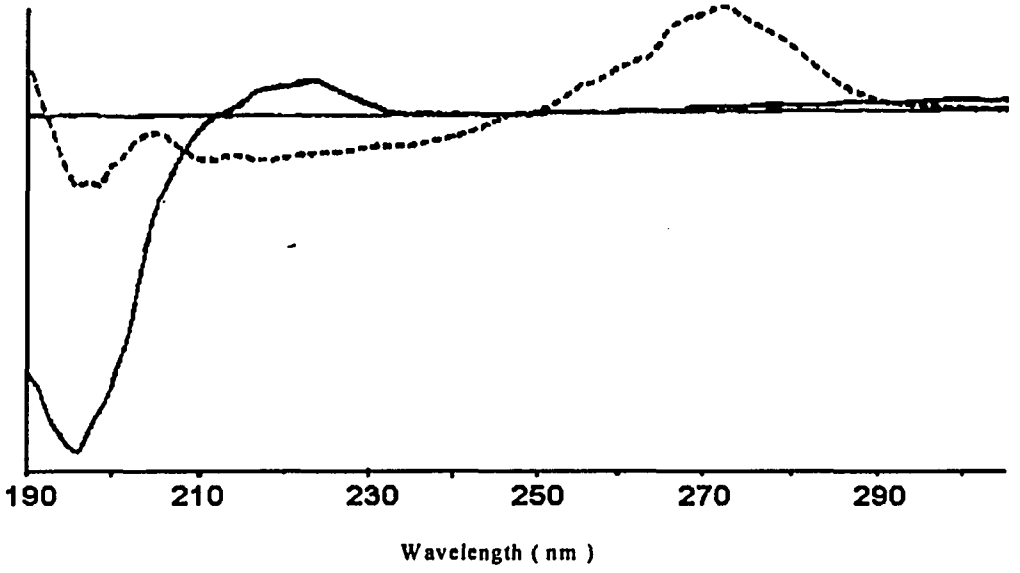


Figure 6-1. CD spectra of ΔTAR alone at 88 μM (- - - -) and Tat-peptide alone at 88 μM (——).

CD spectrum of ΔTAR (dashed line) has a positive maximum around 270 nm, which is associated with the A-conformation of a nucleotide helix (Loret, et al.,1992). The

CD spectrum of Tat-peptide (solid line in Figure 6-1) is characterized by a negative band about 200 nm, which is normally attributed to a

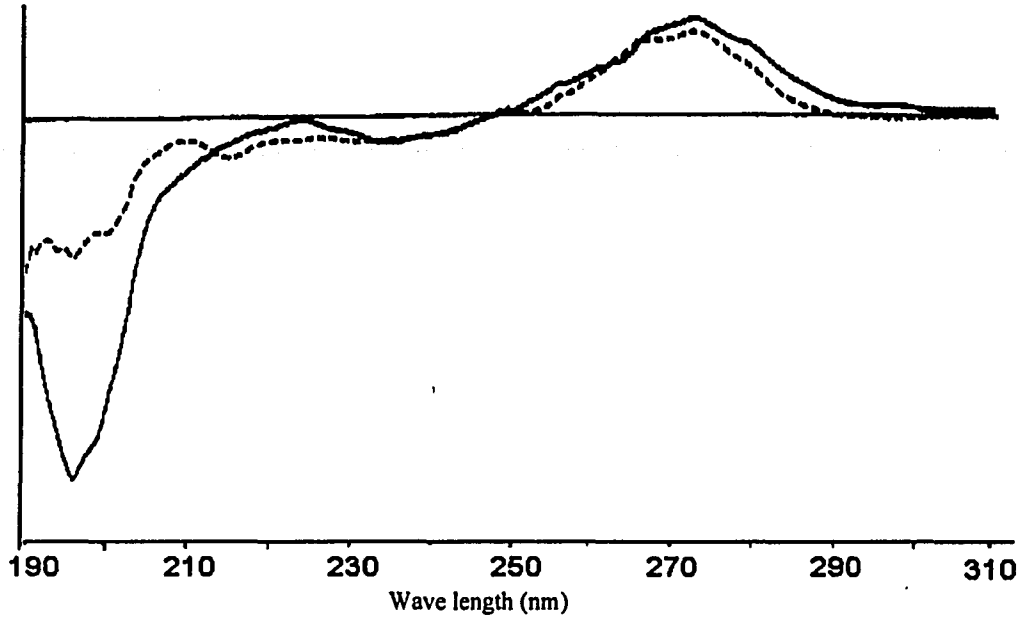


Figure 6-2. CD spectra of the Tat-peptide and TAR 1:1 complex at 88μM(- - -) and the sum of the two spectra in Figure 6-1(——).

random coil structure. The dashed line in Figure 6-2 represents the CD spectrum of Tat-peptide and ΔTAR complex. Compared with the solid line, which is the sum of the two spectra in Figure 6-1, we can find an approximately 15%-20% intensity decrease at about 270 nm. The change was assigned to the RNA because ΔTAR alone shows a large maximum near this wavelength and no known peptide conformation contributes at this wavelength range. The 270 nm band of RNA is sensitive to base stacking. The decrease in Δε upon forming the complex can be interpreted as a

modification of the base stacking induced by the binding of Tat-peptide (Loret, et al., 1992). It was also found that when an A-form polynucleotide is perturbed toward a B-form, the 270 nm band will decrease in intensity with a red shift at the crossover (Goodman, et al., 1969). This is similar to the observation reported in Figure 6-2 for the ΔTAR-Tat-peptide interaction. Therefore, it is suggested that the TAR RNA assumes a more B-like character upon binding with Tat-peptide (Loret, et al., 1992). This A to B form transition can also be noticed in the VCD spectra change of ΔTAR and its binding complex.

There is also a CD signal change around 200 nm upon forming a 1:1 complex. Because of the peptide and the RNA signals overlap in this region of the spectrum, it may consist of contributions from both peptide and RNA conformational changes. It seems likely that the peptide contributes significantly to this signal change because the peptide itself has a negative peak at 190-210 nm and this signal change is much larger than might be expected from changes in RNA alone if we compare it to the 270 nm peak decrease. This leads to a conclusion that the Tat-peptide changes its conformation upon binding with ΔTAR.

6.2 VCD Spectra

As we discussed in the Section 2 of Part I, VCD is exceptionally sensitive to macromolecular conformational change. Here, the VCD spectra of Tat-peptide and

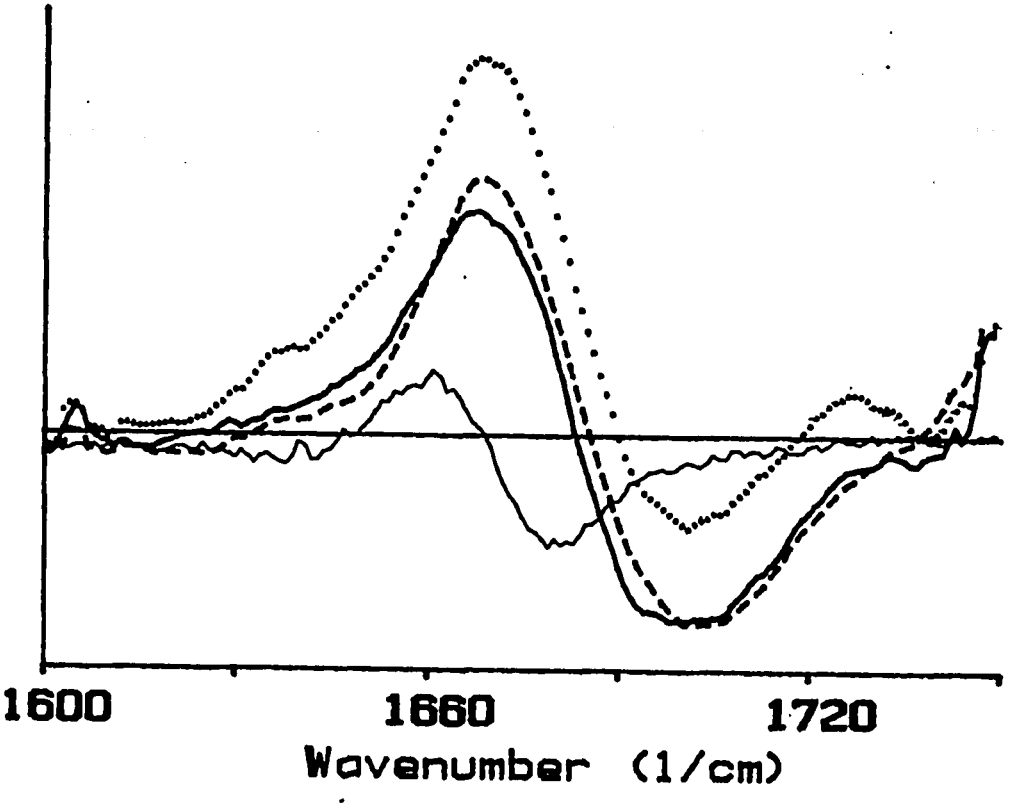


Figure 6-3. VCD spectra of the Tat-peptide alone at 2.5 mM (—), ΔTAR alone at 2.5 mM (...), 2.5 mM ΔTAR + 2.5 mM arginine (—) and 2.5 mM ΔTAR + 2.5 mM Tat-peptide(---).

In Figure 6-3. The dotted line (.....) is the VCD spectrum of ΔTAR alone at 2.5 mM concentration. This non-conservative positive-negative VCD signal is the result of C=O coupling among the bases. Upon binding with an equal molar amount

of Tat-peptide (dashed line in Figure 6-3), the positive non-conservative couplet changed to a positive conservative one with the zero crossing shifting toward lower energy. In the first part of this paper, when we discussed the VCD spectrum of poly(rG).(rC), we noticed that one of the VCD spectral differences between A-form poly(rG).(rC) and B-form poly(dG).(dC) is that A-form's signal (poly(rG).(rC)) is positively biased and the corresponding B-form's signal (poly(dG).(dC)) is much more conservative (Zhong, et al., 1990). The changes of conservativeness in the VCD signal pattern are related with an A-form to B-form transition of the double stranded polynucleotide. This observation confirms the conclusion from CD spectra: upon binding with Tat-peptide, ΔTAR RNA assumes a somewhat more B-like character. Compared to the small CD signal change at 270 nm (Figure 6-1), VCD is much more sensitive to the modification of the base stacking and A- to B-form transition induced by the binding of Tat-peptide.

Competition experiments with arginine and arginine analogs and experiments with an arginine affinity column have demonstrated that TAR contains a specific arginine-binding site and that the guanidine group of arginine is essential for binding TAR (Tao, et al., 1992). CD experiments have suggested that specific binding of Tat-peptide, arginine, or guanidine all induce a conformational change in TAR (Tan, et al., 1992). Here we measured the VCD spectrum of ΔTAR-arginine 1:1 complex at the concentration of 2.5 mM (dark solid line in Figure 6-3) and compared it with the VCD spectrum of ΔTAR-Tat-peptide at the same concentration. The fact that these

two VCD spectra are almost the same indicates that the single arginine and Tat-peptide have a similar effect on the RNA conformation. On the other hand, because a single amino acid will not produce any VCD signal, the spectral difference between ΔTAR alone (.....) and ΔTAR-arginine complex (---) indicates the binding induced conformational change of RNA. The small difference between two spectra, ΔTAR-arginine complex and ΔTAR-Tat-peptide complex, may be considered as belonging to the conformation of the Tat-peptide upon binding with RNA (Figure 6-4).

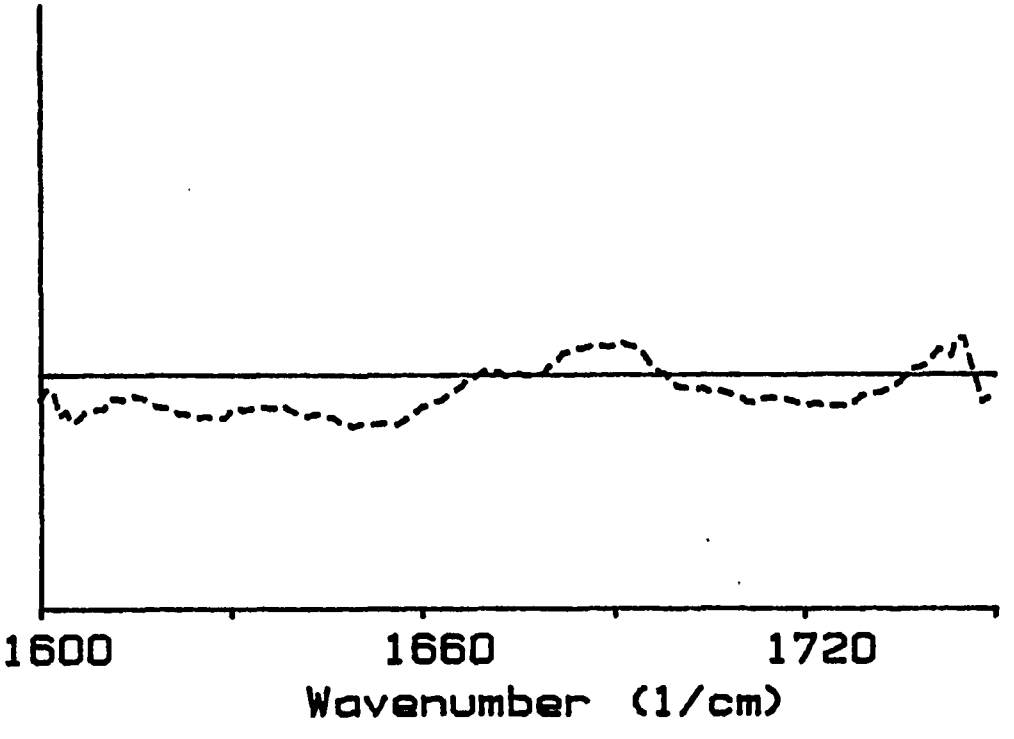


Figure 6-4. Subtraction result of the VCD spectra of the Tat-peptide upon binding to the RNA.

Compared with the VCD spectrum of ΔTAR, Tat-peptide itself (narrow solid line in Figure 6-3) has a much weaker VCD signal in this wavelength range. This weak signal may be due to the short length of the peptide precluding it to form a unique structure with a well defined solution conformation and less C=O coupling than that of nucleotides. Never the less, the negative (1635 cm^{-1}), positive (1660 cm^{-1}) and negative (1678 cm^{-1}) peaks suggest that the Tat-peptide has some secondary structure.

6.3 Prediction of Tat-peptide Conformation

Functions of proteins or peptides arise from their conformations. Primary structure is the sequential order of the residues. Secondary structure describes the helices of the residues. Tertiary structure is the three-dimensional arrangement of residues. Quaternary structure is the arrangement of subunits. Helices of the residues is due to the rigidness of the peptide units which are the building blocks of proteins or peptides.

In a peptide, amino acids are joined by a peptide linkage formed by the α-carboxyl group of the first amino acid with the α-amino group of the second amino acid release of a water to form a peptide bond. Two amino acids form a dipeptide by this acid-amide linkage. This process is repeated and peptides are formed from a

continual sequence of peptide bonds and α-C-atoms (C_α). A characteristic amino acid side chain R is attached in every C_α.

The peptide bond exhibits a resonance-stabilized structure (Figure 6-5). In the free rotation form, the free electron pair of nitrogen is localized in a hybridized

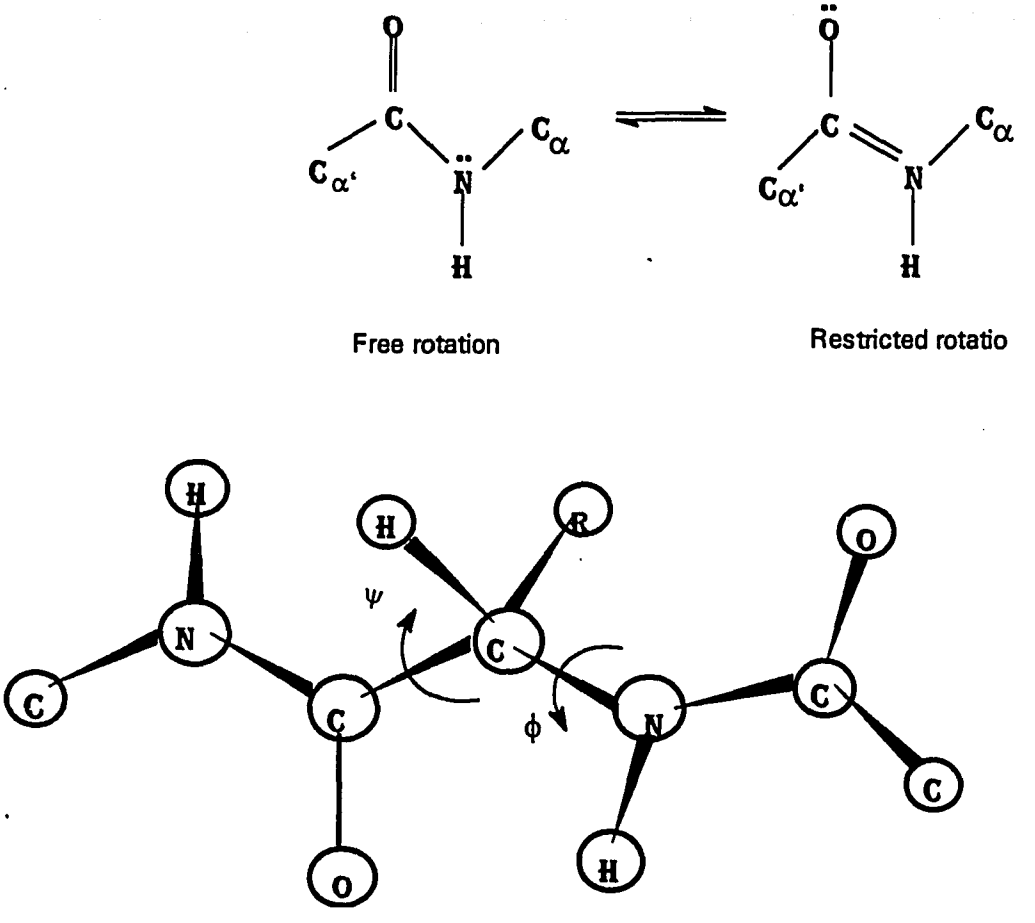


Figure 6-5. Resonance structure and dimensions of the peptide bond

2p-orbital. In the restricted rotation form, the single electron pair is in the 2p-orbital of the O atom. Due to the electron shift from N to O, the N atom is positively charged and the O atom is negatively charged. This resonance structure results in delocalized π electrons and causes the molecular orbitals of all three atoms, O, C, and N, to be stretched. The C-O bond loses its double-bond characteristics by an increase in the C-O bond length. The C-N bond gains a partial double-bond character by a shortening in the C-N bond length. Therefore, the peptide unit becomes rigid. The hydrogen of the substituted amino group is nearly always trans (opposite) to the oxygen of the carbonyl group. There is no freedom of rotation about the bond between the carbonyl carbon atom and nitrogen atom. The six atoms C_α-CONH-C_α lie in a plane (Figure 6-5). The only degrees of freedom they have is to rotate around two single bonds: C_α-C bond and N-C_α bond. The dihedral angles are represented by ϕ (N-C_α) and ψ (C_α-C) The secondary conformation of a peptide chain can be fully described by the angles ϕ and ψ. The potential energy of the residues plotted against ϕ and ψ is called a Ramachandran plot, which describes the possible range of ϕ and ψ for a pair of peptide bonds (Carton and Schimmel, 1980). The process of folding of a peptide chain is not understood in detail at present. However, it is known that it occurs spontaneously.

The secondary structure of Tat-peptide in solution has been investigated. Molecular models of this arginine rich region of Tat resulted from CD studies suggest that this basic region is an extended structure (Loret, et al., 1991) or is unstructured in

the absence of RNA (Calnan, et al., 1991b). CD spectrum (Loret, et al., 1992) demonstrated that with an increasing the peptide length to 47-72 amino acids (see Figure 4-2), α-helix structure increased at the residues of 59-72 and the 47-58 region became an extended structure. The arginine-rich 49-58 region is primarily responsible for electrostatic interaction with the phosphates of RNA and the arginine side chain can additionally interacts with substituent groups of the nucleotide bases to confer base recognition in the complex.

Another group (Mujeeb, et al 1994) studied a 25- amino acid hybrid peptide that contained the basic domain of Tat-protein (48-57) fused to a 15-residue activation domain from the Tat protein of equine infectious anemia virus (EIAV, another retrovirus). Their CD and NMR results indicated that Tat basic domain formed an α-helix, whereas the adjacent regulatory sequence is mostly in an extended conformation. They suggested that the tendency to form a helix may be a common property of arginine- and lysine- rich RNA-binding domains.

Although the CD spectrum of Tat-peptide basic domain (48-57) showed a random coil pattern, which suggested that this arginine-rich region of Tat is unstructured in the absence of RNA (Calnan, et al., 1991), we believe this short linear peptide is not truly unordered because a truly unordered peptide would lack a VCD spectrum. The VCD spectrum of Tat-peptide has a negative peak at 1635 cm^{-1}, a positive peak at 1660 cm^{-1} and another negative peak at 1678 cm^{-1}. The spectrum stays the same when the concentration increases from 1.4 mg/ml to 14 mg/ml (Figure

6-6) and the temperature changes from $10^{o}C$ to $20^{o}C$ (data not shown). If the temperature is increased to $40^{o}C$, the magnitude of the 1678 cm^{-1} peak decreases probably due to the partial melting of the peptide structure (Figure 6-7). Upon addition of SDS (sodium dodecyl sulfate), which can destroy the secondary structure of peptides, the Tat peptide shows a decrease in the magnitude of the VCD negative peak at 1678 cm^{-1} (Figure 6-8), the CD positive peak at 220 nm and the CD negative

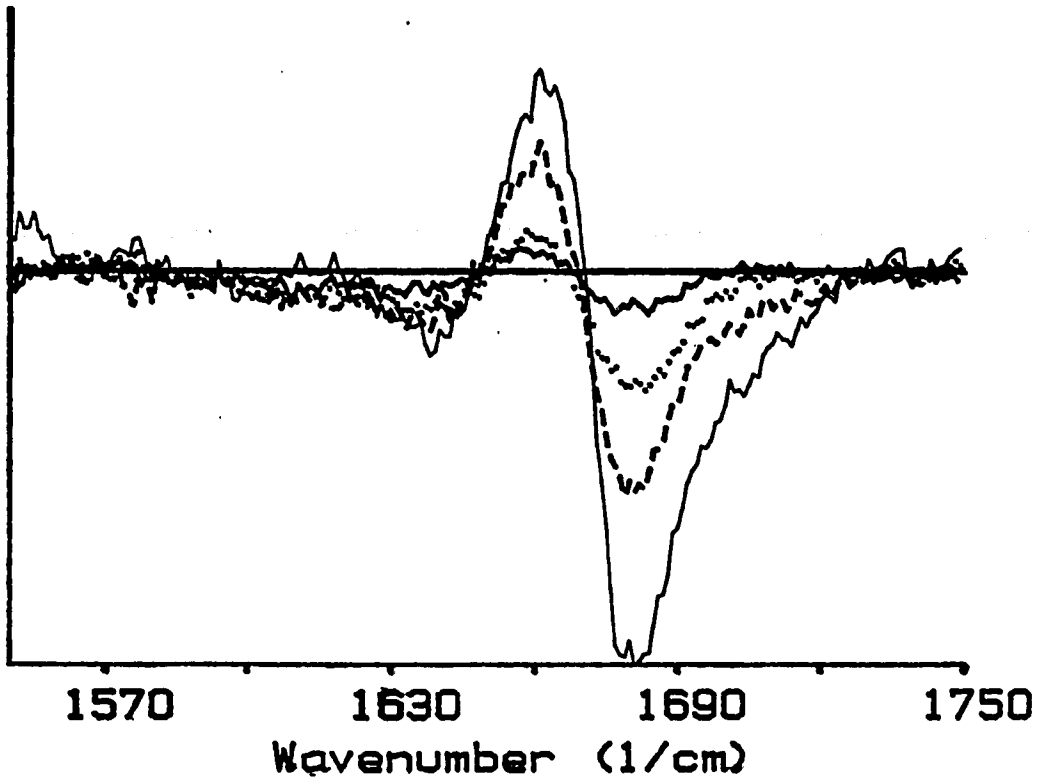


Figure 6-6. VCD spectra of Tat-peptide in the concentrations of 1.4 mg/ml (—), 2.8 mg/ml (...), 7 mg/ml (---) and 14 mg/ml(—).

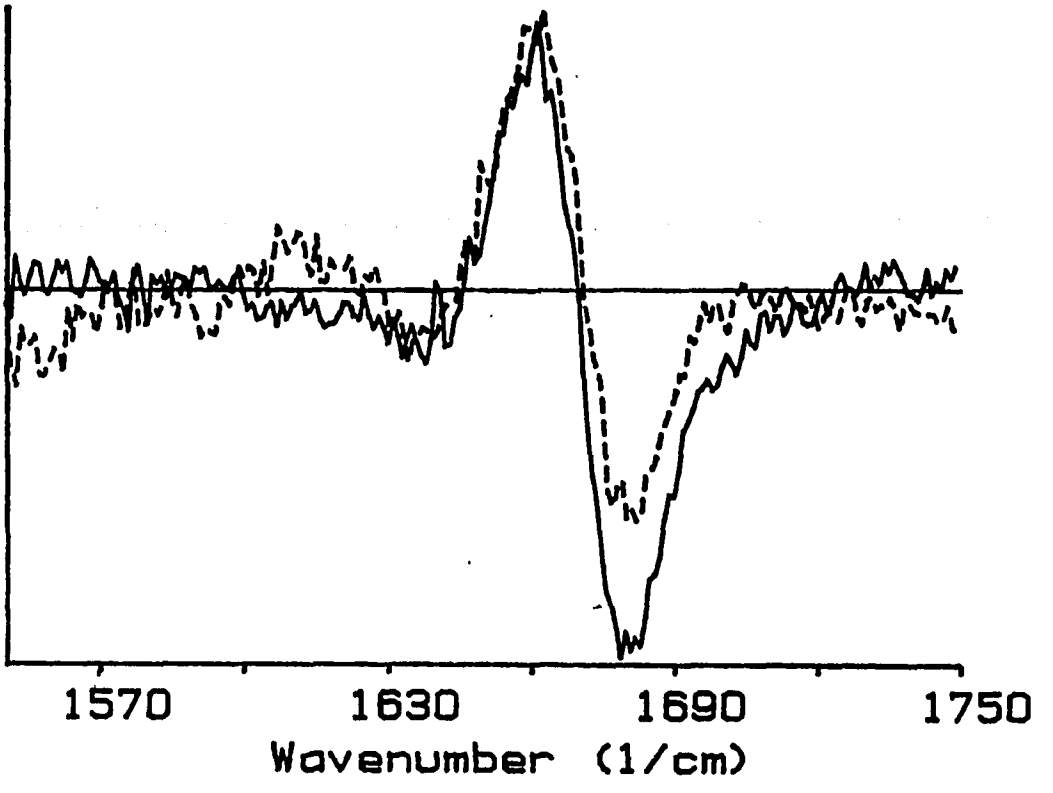


Figure 6-7. VCD spectra of Tat-peptide in 7mg/ml at 10^{o} C (—) and 40^{o}C (---).

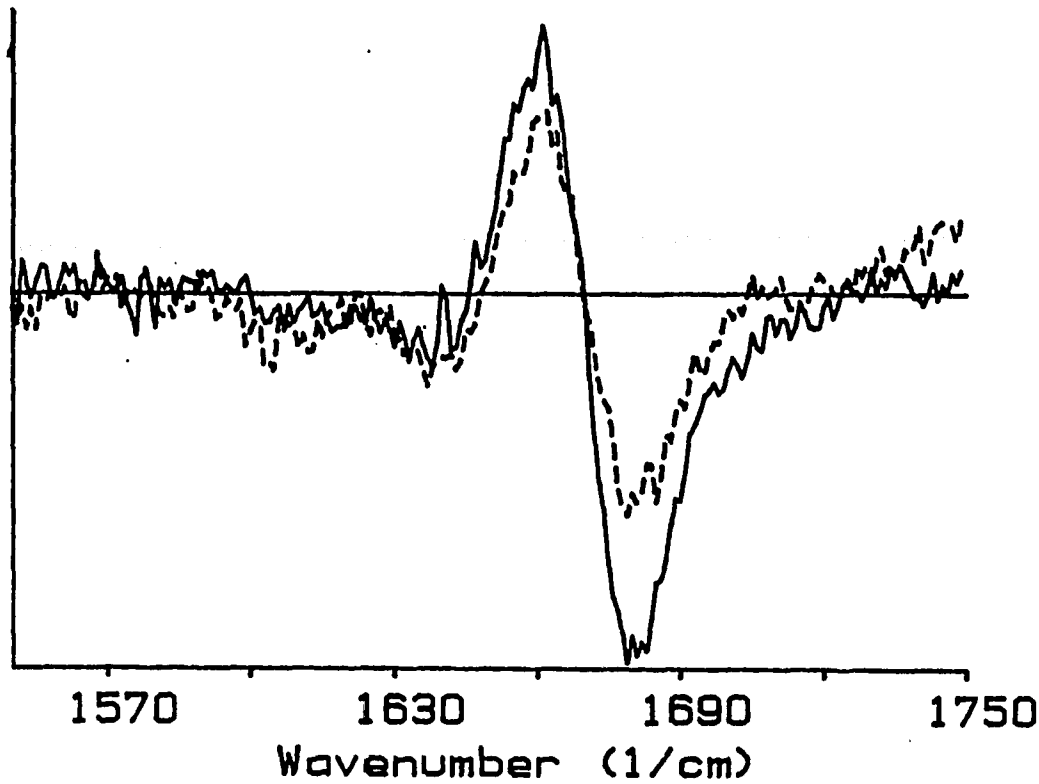


Figure 6-8. VCD spectra of Tat-peptide in 7mg/ml alone (—) and with 8 mM SDS (- - -).

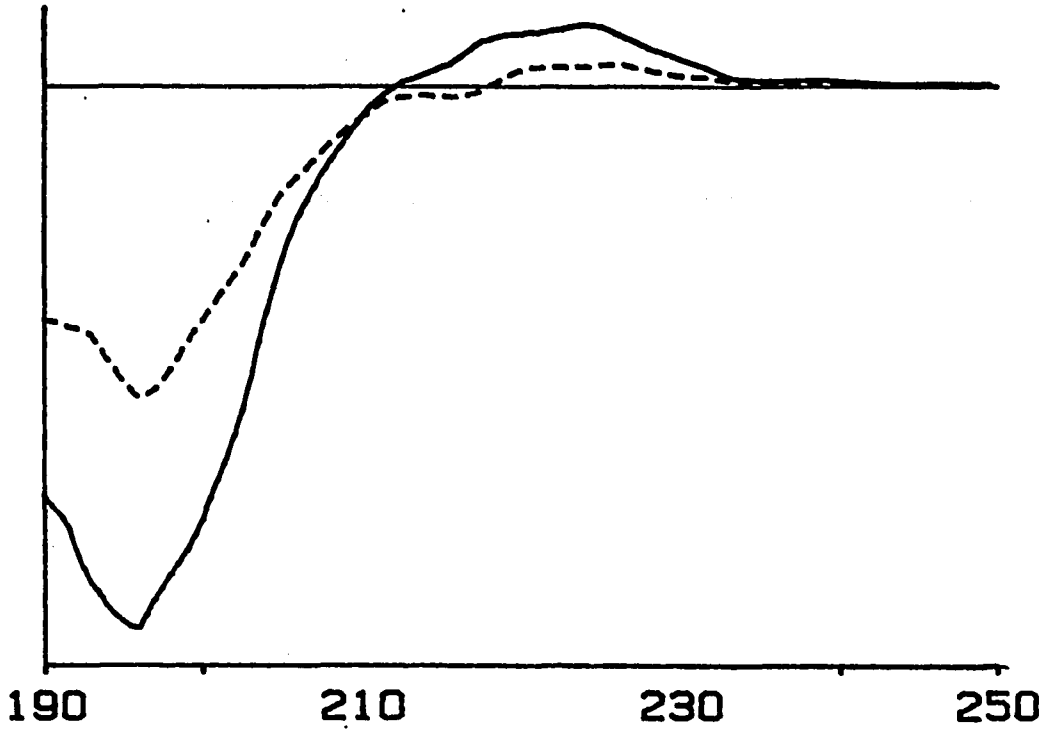


Figure 6-9. CD spectra of Tat-peptide in 1.4mg/ml alone (—) and with 4mM SDS (---).

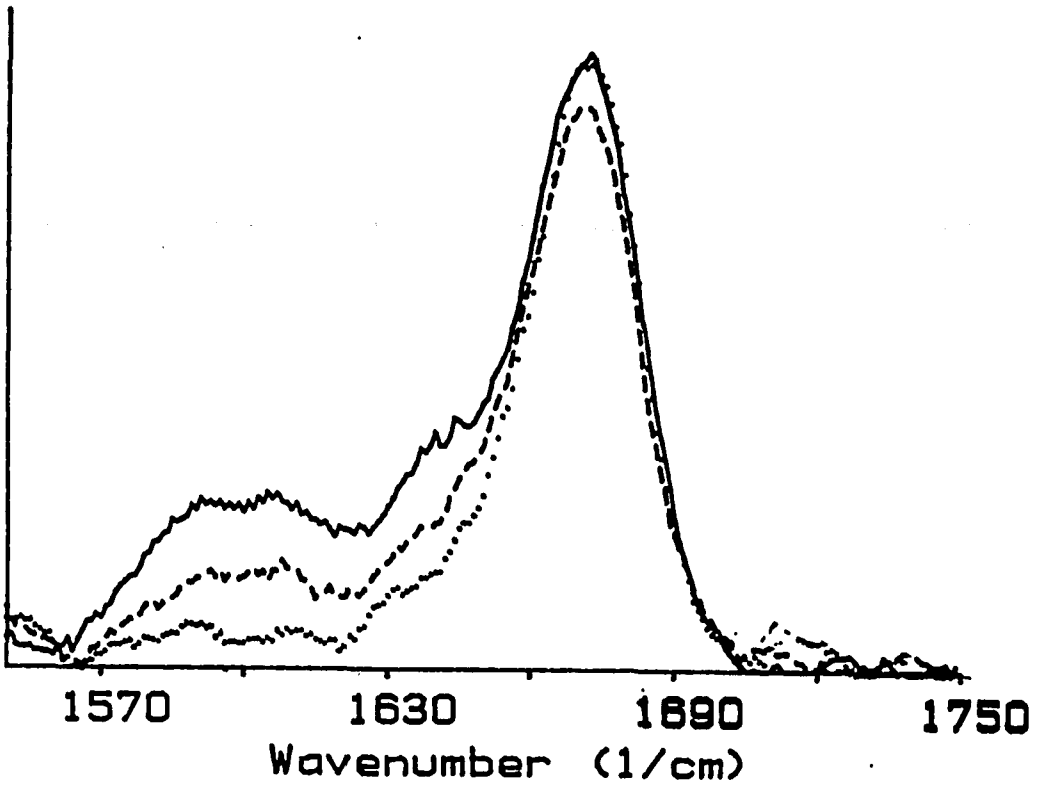


Figure 6-10. IR spectra of Tat-peptide in 2.8 mg/ml alone (—) and with 4 mM SDS (- --) and with 8 mM SDS (...)

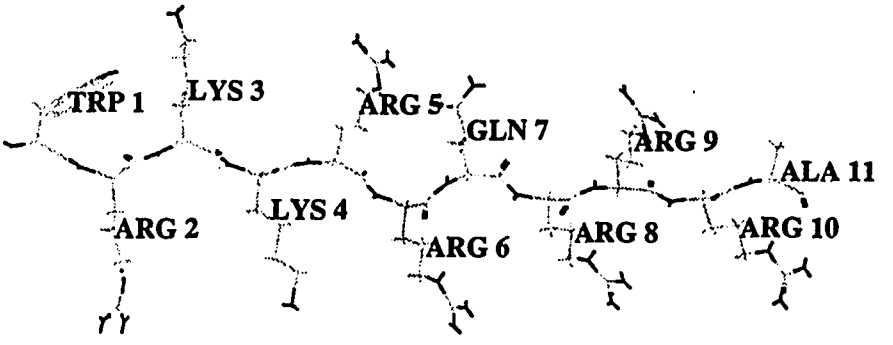
peak at 195 nm (Figure 6-9). The IR spectra also shows a decrease in the magnitude of the 1631 cm^{-1} peak when the concentration of SDS is increased (Figure 6-10).

These data suggest the Tat-peptide is not truly unordered. The VCD pattern implies that a substantial local ordering exists in this " random-coil".

The conformational sensitivity of VCD in the amide I region of the peptide spectrum is attributed to the dipole coupling of the C=O stretching vibrations of the peptide linkages (Diem, 1993). The coupling of these transitions produces distinct VCD couplets which may permit a quantitative determination of dihedral angles between the coupled oscillators and thus the molecular solution conformation. Based on the exciton theory and DECO model which we have introduced in Section 2 of Part I and considering the coupling of the carbonyl groups in the peptide units which are arranged in a fixed geometry, the computation of rotational strength for the Tat-peptide was carried out using Cartesian coordinates of the C and O atoms of the carbonyl groups from the geometry optimization results by the HyperChem. Our calculational results suggested the Tat-peptide is probably an extended left-handed helix with a slightly higher twist than β-sheet. Figure 6-11 shows the geometry and dihedral angles of the Tat-peptide as determined by these calculations. Figure 6-12 is the comparison of experimental VCD spectrum with the calculated spectrum in which the geometry in Figure 6-11 was used.

Extended helices seem to be less well defined than, for example, α-helices, and appear in a broader minimum of the Ramachandra plot. Their helicity is very low: 2.2 - 2.4 residues per turn. For comparison: 2.0 residues per turn is a β-sheet conformation with $\phi = -120^{o}$ and $\varphi = 113^{o}$, 3.6 residues is an α-helix conformation.

Thus, the 2.2 residues corresponds to a "sheet with a twist" structure. In terms of conformational angles, these structures are very similar to a parallel sheet. In previous



ϕ	-121.7	-138.5	-140.5	-131.7	-133.2	-125.2	-136.0	-140.5	-137.0	-125.0
ψ	141.6	133.8	122.4	136.0	136.1	105.6	119.0	136.7	153.6	158.0

Figure 6-11. Tat-peptide, extended left-handed helix with a slightly higher twist than β sheet.

VCD calculations, it was found that between very similar structures, for example the 2.2 and 2.4 residues/turn conformations, the sign of the VCD switches: for 2.4 residues, the sign was similar to the right-handed form; for 2.2 residues, the sign was typical of the left handed form as in the Tat-peptide.

The left-handed helix conformation has been found for other short peptides by VCD spectra (Diem, 1993). The CD of these peptides all showed a "random coil", but the VCD features calculated for such an extended helix reproduced the observed data well. All of these demonstrate that VCD is a powerful solution conformation probe. The combination of CD and VCD spectral analysis can give us a more information and a better understanding of the solution conformation of macromolecules.

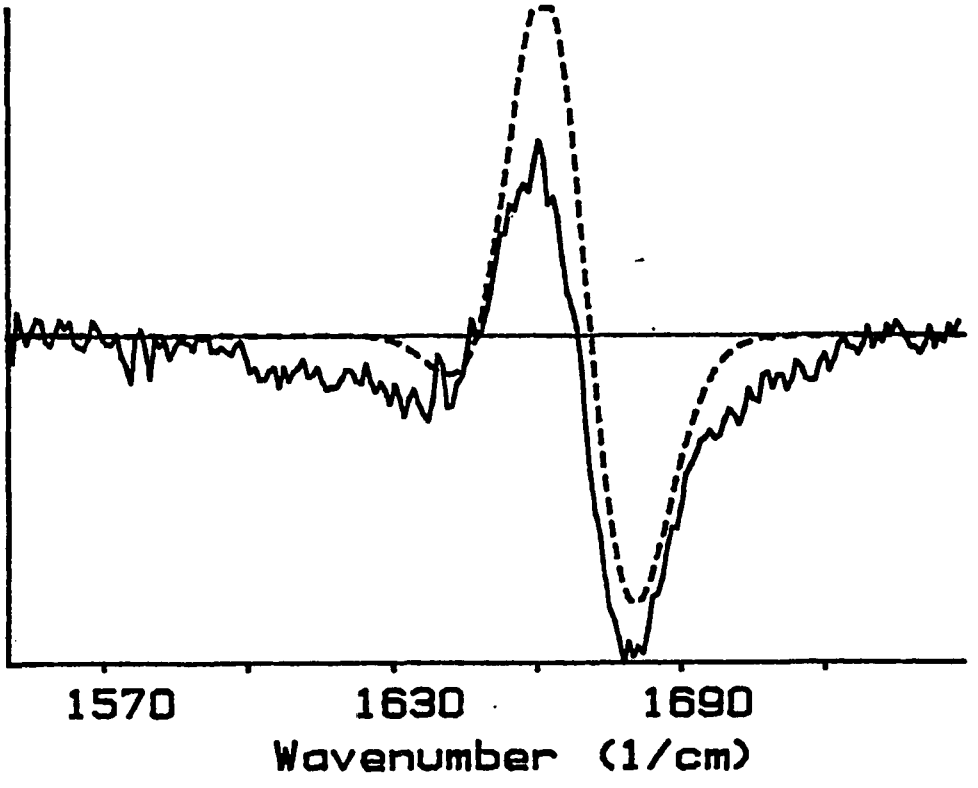


Figure 6-12. Comparison of the calculated VCD spectrum with the experimental VCD spectrum for the Tat-peptide. Calculated VCD(---), experimental VCD (—).

6.4 Molecular Modeling of Tat-TAR Interaction

RNA structure is very important in Tat-TAR recognition. Binding, mutagenesis and chemical modification data strongly support the suggestion that the identical set of nucleotides -U^{23}, G^{26}, C^{39}, A^{27}, U^{38}- is important for Tat transactivation, Tat binding, Tat-peptide binding and arginine binding (Puglisi, et al., 1993).

The conformations of TAR with an arginine analog specifically bound at the binding site for Tat were characterized by nuclear magnetic resonance (NMR) spectroscopy (Puglisi, et al., 1992; 1993). Upon arginine binding, the bulge changes conformation, and essential nucleotides for binding, U^{23} and A^{27}-U^{38} form a base-triple interaction that stabilizes arginine hydrogen bonding to G^{26} and phosphate. Figure 6-13 is their schematic representation of the interaction of arginine guanidinium group with G^{26} and two phosphate groups.

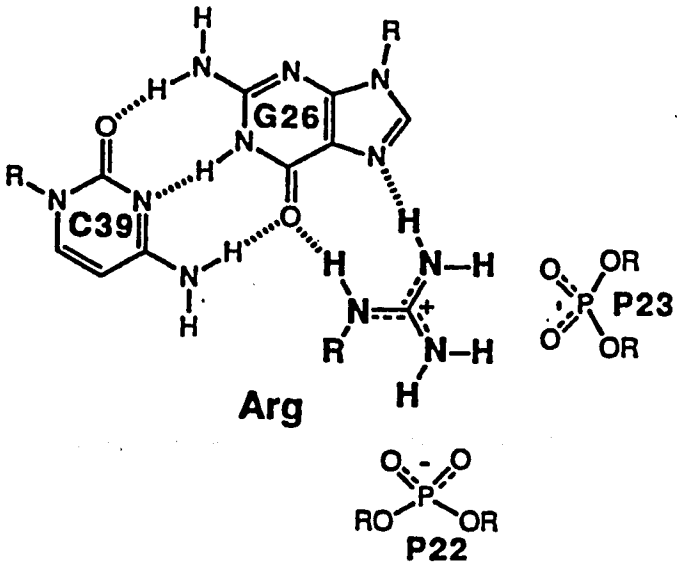


Figure 6-13. Schematic representation of the interaction of an arginine guanidinium group with TAR.(Puglisi, et. al., 1993)

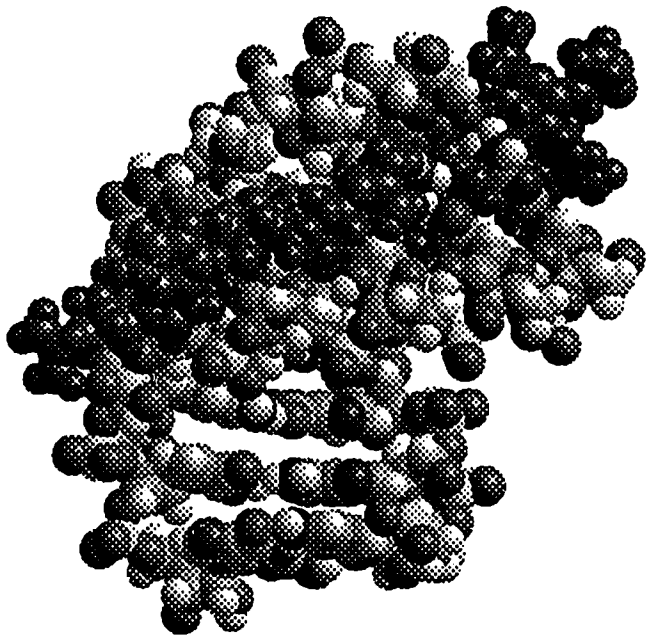


Figure 6-14. Molecular modeling result of the binding of Tat-peptide and ΔTAR RNA.

In order to locate the most favorable position of these two molecules, energy minimization for the interaction of ΔTAR with Tat-peptide was performed with the use of HyperChem's molecular mechanics optimization with the AMBER (Assisted Model Building and Energy Refinement) force field, Polak-Ribiere optimization algorithm, distance dependent dielectric. The elecctrostatic and van der Waals scale factors were set to 0.5 and no addition charge was considered. As a starting point, the RNA structure which was provided by J. Williamson and the Tat-peptide structure in Figure 6-11 were used. Arginine 5 (see Figure 6-10 or #52 in Figure 4-2) was positioned to interact with G^{26} and two phosphate groups as suggested by NMR data (Figure 6-13). The result shows that at the RNA loop region, G^{26}, A^{27}, U^{38},C^{39},U^{23} are required for the Tat-peptide specific binding which is consistent with the result determined by the sequence variation and gel electrophoretic partition methodology (Weeks, et al., 1991). For the Tat-peptide, besides arginine 5 which interacts with G^{26} and two phosphate groups, the arginine 6 (or #53) arginine 8 (or #55), and lysine 4 (or #51) also form H-bonds with the phosphate groups at the RNA loop region. Figure 6-14 shows a model of the binding of these two molecules.

6.5 Kinetics

The dissociation rate (k_{off}) has been studied for the reactions of ΔTAR with Tat-peptide (14 to 38-residue peptides derived from Tat that spans the basic region,

Weeks, et al., 1991 and 1992; Long & Crothers, 1995) using a gel electrophoretic method, the reaction was studied by incubating radiolabeled RNA with peptides, then challenging the complex with an excess of unlabeled competitor, and assaying for the amount of remaining radiolabeled complex at appropriate time intervals. k_{off} was 0.11 to 0.41 min^{-1} for Tfr38 (38-residue peptide derived from Tat that spans the basic region) and 0.075 to 0.061 min^{-1} for Tfr24. Association rate for ΔTAR- Tfr24 reaction was measured in a similar method but using a quenching protocol (Long & Crothers, 1995). The calculated association rate constant was at the level of 1×10^{8} M^{-1} s^{-1}.

In order to study the mechanism of the forward reaction, stopped-flow spectrofluorimetry was used to evaluate the elementary steps involved in the reaction since the fluorescence of tryptophan at the N-terminus of the Tat-peptide decreases upon binding with ΔTAR. The observed rate constant, $k_{obs,}$ for the change in fluorescence is related in a concentration-dependent manner on the type of elementary steps to which the mechanism belongs (Gupta, et al.,1992). For example, for a single bimolecular reaction between RNA [R] and peptide [P], the observed first order change in fluorescence is related to the association and dissociation rate constants and excess component [R] as following equations:

$$R + P \underset{k_{-1}}{\overset{k_1}{\Leftrightarrow}} RP \qquad (6-1)$$

$$k_{obs} = k_{-1} + k_1[R]_0 \qquad (6-2)$$

Here $[R]_0$ is the initial RNA concentration. The second possibility is that of RNA undergoing slow transformation between two states of which only R^* is able to associate with the peptide.

$$R \underset{k_{-1}}{\overset{k_1}{\Leftrightarrow}} R^* + P \underset{k_{-2}}{\overset{k_2}{\Leftrightarrow}} R^* P \qquad (6-3)$$

$$k_{obs} = k_1 + k_{-1} \frac{K_{-2}}{K_{-2} + [R]_0} \qquad (6-4)$$

$$where \quad K_{-2} = \frac{k_{-2}}{k_2}$$

The last case involves the rapid formation of an intermediate, RP, which isomerizes to yield the final complex RP^*.

$$R + P \underset{k_{-1}}{\overset{k_1}{\Leftrightarrow}} RP \underset{k_{-2}}{\overset{k_2}{\Leftrightarrow}} RP^* \qquad (6-5)$$

With the assumptions: (i) pseudo-first-order approximation, (ii) steady state for [RP], (iii) $k_{-2} << k_2$, k_{obs} is related to various rate constants and $[R]_0$ by

$$\frac{1}{k_{obs}} = \frac{1}{k_2} + \frac{k_{-1} + k_2}{k_1[R]_0 k_2} \qquad (6\text{-}6)$$

Since k_{obs} depends in a characteristic manner on $[R]_0$, it could be exploited to discriminate experimentally between different mechanisms. From Equation (6-2), k_{obs} would increase linearly with $[R]_0$; Equation (6-6) predicts that $1/k_{obs}$ would increase linearly with $1/[R]_0$; while according to Equation (6-4), k_{obs} would decrease as $[R]_0$ increases.

For our system, the observed first order rate constant, k_{obs}, was evaluated from the slopes of linear plots of $\ln(F_\infty - F_t)$ versus time(t), where F_∞ and F_t are the fluorescence intensity at infinite time and time t, respectively (Figure 6-15). A plot of k_{obs}^{-1} with $[\Delta TAR]^{-1}$ shows a straight line (Figure 6-16), which is consistent with the third mechanism. The intercept equals to $1/k_2$ and the slope equals to $\frac{k_{-1} + k_2}{k_1 k_2}$. From the intercept of $k_{obs}^{-1} \sim [\Delta TAR]^{-1}$ in Figure 6-16, which is 0.0144 sec, $k_2 = 69.4\ sec^{-1}$. With an additional assumption of $k_{-1} << k_2$, the slope $\approx 1/k_1$. From the slope of $k_{obs}^{-1} \sim [\Delta TAR]^{-1}$ in Figure 6-16, which is 0.0146 μM sec, $k_1 = 0.7\times10^8\ M^{-1} sec^{-1}$. This result is on the same order of the association rate constant $1\times10^8\ M^{-1} sec^{-1}$ derived from the quenching protocol (Long & Crothers, 1995).

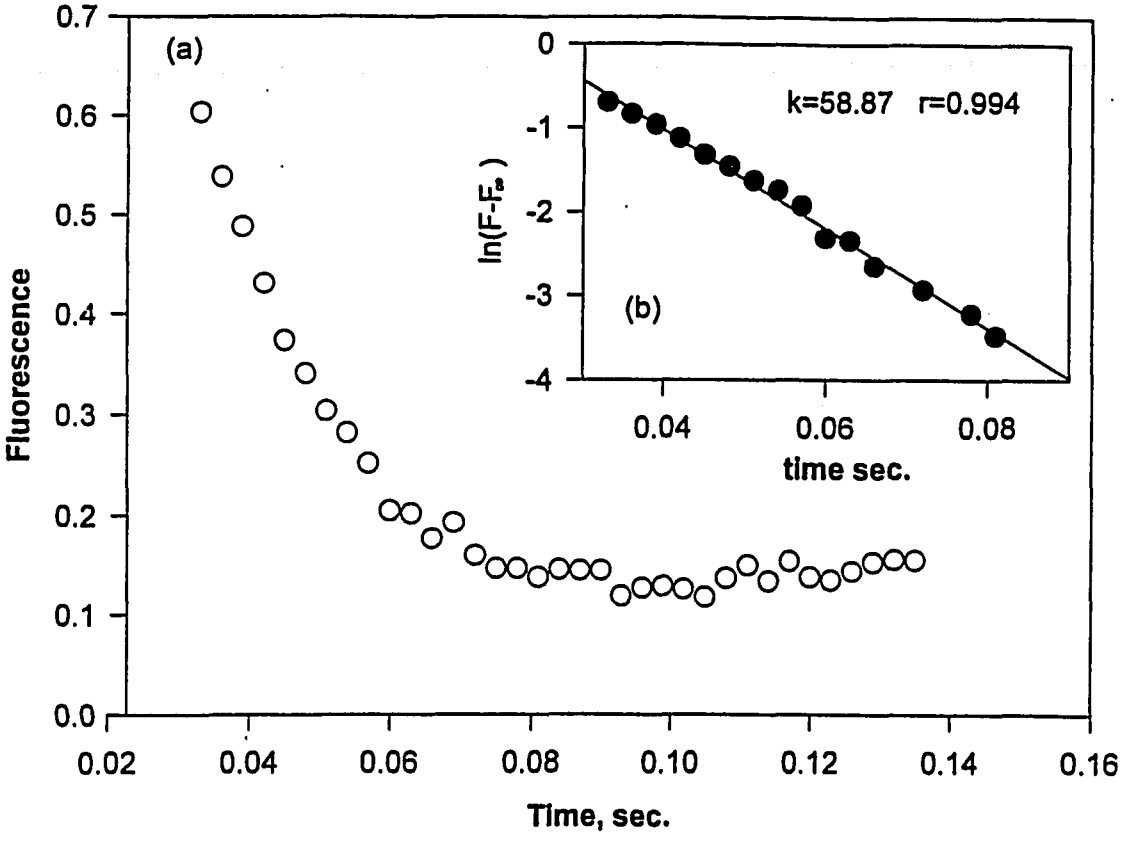


Figure 6-15. Example of a stopped flow fluorescence trace for the reaction of Tat-peptide with ΔTAR RNA.

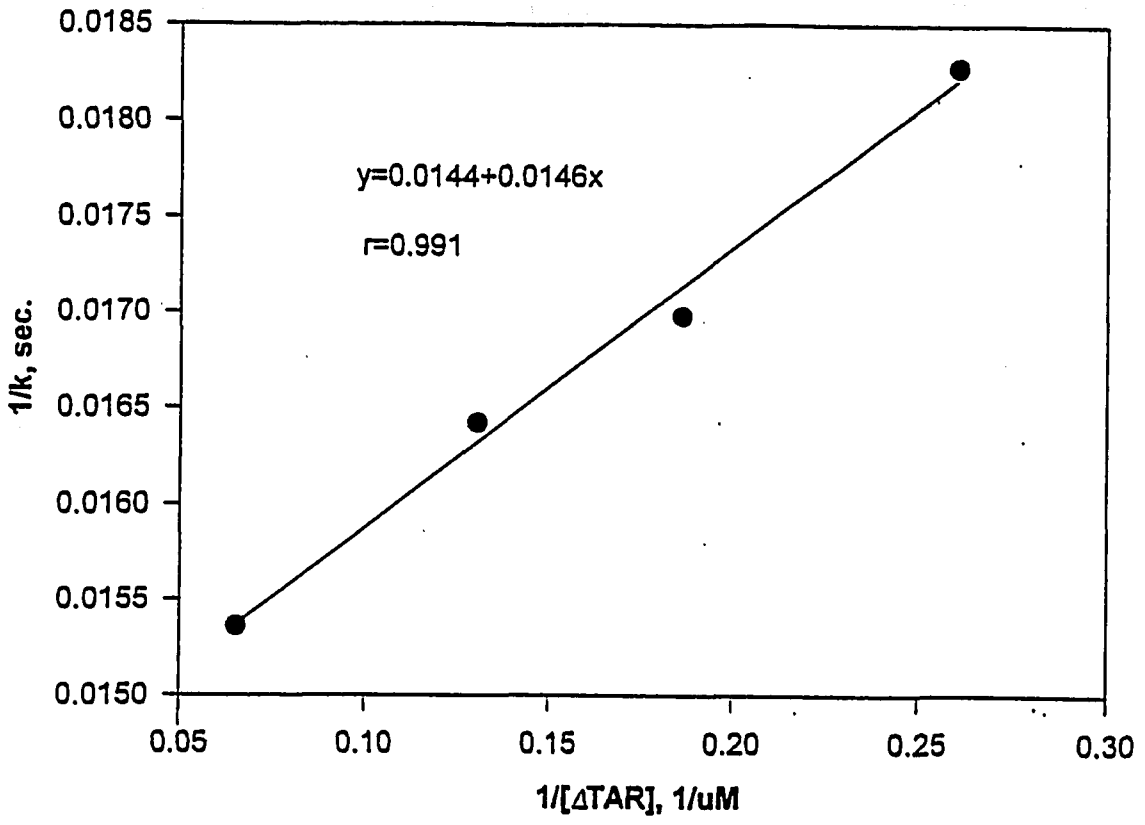


Figure 6-16. Determination of rate constants for the reaction of Tat-peptide with ΔTAR. The ordinate of this plot gives k_2.

6.6 Summary

Human immunodeficiency virus type I (HIV-I) encodes a trans-activating regulation protein, called Tat, that is needed for efficient transcription of the viral genome. HIV-I Tat acts by binding a RNA stem-loop structure which is the trans-activation response element (TAR). Here we present the spectroscopic studies of Tat-peptide, which is the basic domain of the Tat-protein, and ΔTAR, which is the shortened form and the effective part of TAR, interaction.

We measured the CD spectra of ΔTAR, Tat-peptide and their 1:1 complex. The results are consistent with the results from other group. The conformational change in the ΔTAR and Tat-peptide can be identified by comparison of the spectra before and after complex formation.

We also measured the VCD spectra of ΔTAR, Tat-peptide, ΔTAR-Tat-peptide 1:1 complex and ΔTAR-arginine 1:1 complex. ΔTAR changes its VCD pattern from an no-nconservative to a conservative couplet upon binding with Tat-peptide or arginine. Single arginine and Tat-peptide have a similar effect on the ΔTAR conformation. Binding induced conformational changes can be identified by the comparison between spectra.

The VCD of Tat-peptide, which has a random coil CD pattern, shows a negative peak at 1635 cm^{-1}, a positive peak at 1660 cm^{-1} and another negative peak at 1678 cm^{-1}. The spectrum stays the same with changes of concentration and low temperature but changes with high temperature and high polar solvent. Based on the

exciton theory and DECO model, we performed VCD calculations to predict the possible solution structure of this peptide. The conformational sensitivity of VCD in the amide I region of the peptide is believed to be due to the dipole coupling of the C=O stretching vibrations of the peptide linkages. The coupling of these transitions produces distinct VCD couplets which may permit a quantitative determination of dihedral angles between the coupled oscillators and thus the molecular solution conformation. The calculated VCD spectrum using an extended left-handed helix conformation can reproduce the observed spectrum well. Using this extended left-handed helix conformation, we performed molecular modeling to simulate the binding of Tat-peptide to the loop region of ΔTAR. The result confirms that G^{26}, A^{27}, U^{38},C^{39},U^{23} are required for the Tat-peptide specific binding.

Stopped-flow spectrofluorimetry was used to investigate the mechanism of the forward reactions of this system. The fluorescence of tryptophan at the N-terminus of Tat-peptide decreases with time upon the rapid binding with ΔTAR. The observed rate constant, k_{obs} for the change in fluorescence is related in a concentration-dependent manner to a two-step mechanism. $k_1 = 0.7\times10^8\ M^{-1}\,sec^{-1}$ and $k_2 = 69.4\ sec^{-1}$ are obtained from the linear relationship between k_{obs}^{-1} and $[\Delta TAR]^{-1}$.

BIBLIOGRAPHY

PART I

Annamalai, A., Keiderling, T. A. (1987) J. Am. Chem. Soc., **109,** 3125-3132

Arnott, S., Hukins, D. W. L. (1972) Biochem. Biophys. Reserch Comm., **48** 1392-1399

Arnott, S., Chandarasekaran, R., Martila, C. M. (1974) Bioche. J., **141,** 537-543.

Arnott, S., Chandarasekaran, R., Heslie, A. G. W. (1976) J. Mol. Biol., **106,** 735-748.

Bayley, P. M., Nielsen, E. B. Shellman, J. A. (1969) J. Phys. Chem., **73,** 228-243.

Bloomfield, V. A., Crothers, D. M., Tinoco, I. (1974) Physical Chemistry of Nucleic Acids, Harper & Row Publishers, New York.

Bush, C. A. (1973) Physico-chemical Properties of Nucleic Acids, **2,** 147, Academic Press, London, New York.

Cantor, C. R., Schimmel, P. R. (1980) Biophysical Chemistry Part I: The Conformation of Biological Macromolecules, W. H. Freeman and Company,San Francisco.

Diem, M. (1993) Introduction to Modern Vibrational Spectroscopy, John Wiley & Sons., New York.

Fasman, G. D., Lindblow, C., Grossman, L. (1964) Biochemistry, **3(8),** 1015-1021

Freedman, T. B., Nafle, L. A. (1983) J. Chem. Phys., **78,** 27; **79,** 1104.

Gulotta, M., Goss, D. J., Diem, M. (1989) Biopolymers, **28,** 2047-2058.

Hartman, K. A., Lord, R. C., Thomas Jr., G. J. (1973) Physico-chemical Properties of Nucleic Acids, **2, 1,** Academic Press, London, New York.

Holzwarth, G., Hsu, E. C. Mosher, H. S. Faulkner, T. R. Moscowitz, A. (1974) J. Am. Chem. Soc., **96,** 251.

Holzmarth, G., Chabay, I. (1972) J. Chem. Phys., **57,** 1632-1635.

Hopper, W., Lohmann, W. Markl, H.. Ziegler, H. (1983) Biophysics, Springer-Verlag, New York

Hsu, E. C. and Holzwarth, G. (1973). J. Chem. Phys., **59,** 4678.

Lee, O. and Diem, M. (1992) Analy. Instru., **20(1),** 23-43.

Moffit, W. (1961) J. Chem. Phys., **25,** 467-478.

Pancoska, P., Yasui, S. C. Keiderling, T. A. (1991) Biochemistry, **30,** 429-437.

Polavarapu, P. L. (1983). Mol. Phys., **49,** 645.

Nafie, L. A., Walnut, T. H. (1977) Chem. Phys. Lett., **49,** 441.

Rhodes, W. (1961) J. Am. Chem. Soc., **83,** 3609-3617.

Saenger, W.(1984) Principle of Nucleic Acid Structure, Springer-Verlag, New York.

Schellman, J. A. (1973) J. Chem. Phys., **58,** 2882.

Souleil, C., Panijel, J. (1968) Biochemistry, **7,** 7-13.

Still, C. 1989. MacroModel Software. Version 2.5, Columbia University, New York.

Stryer, L. (1988) Biochemistry, W. H. Freeman and Company, New York

Sugeta, H., Marcott, C., Faulkner, T., Moscowitz, R. (1976) A. Chem. Phys. Lett., **40** , 397.

Tinoco, I. (1963) Radiat. Res., **20,** 133-139.

Tsuboi, M., S. Takahashi, Harada, I. (1973) Physico-chemical Properties of Nucleic Acids, **2,** 92, Academic Press, London, New York.

Williams, A. L., Moore, D. (1983) Biopolymers, **22,** 755-786.

Xie, P.,Diem, V. (1995) J. Amer. Chem. Soc., **117,** 429-437.

Yasui, S. C. and T. A. Keidering (1987) Biopolymers, **26,** 1407-1412

Zhong, W., Gulotta, M., Goss, D. J., Diem, M. (1990) Biochemistry, **29,** 7485-7491.

PART II

Calnan, B. J., Biancalana, S., Hudson, D., Frankel, A. D. (1991) Science, **252,** 1167-1171.

Calnan, B. J., Biancalana, S., Hudson, D., Frankel, A.D. (1991b) Genes & Development, **5,** 201-210.

Cantor, C. R., Schimmel, P. R. (1980). Biophysical Chemistry Part I: The Conformation of Biological Macromolecules, W. H. Freeman and Company,San Francisco.

Cordingley, M. G. (1990) Proc. Natl. Acad. Sci. USA, **87,** 8985-8989.

Dingwall, C., Ernbery, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M., Skinner, M. A. (1990) EMBOJ, **9,** 4145-4153.

Diem, M., 1993. Introduction to Modern Vibrational Spectroscopy, John Wiley & Sons, New York

Felberg, B. K., Cladaras, M. H., Cladaras, V., Copeland, T., Pavlakis, G. N. (1989) Proc. Natl. Acad. Sci. USA, **86,** 1495-1499.

Goodman, M., Verdini, A. S., Toniolo, C., Phillips, W. D., Bovey, F. A. (1969) Proc. Natl. Acad. Sci. USA, **64,** 444-450.

Gupta, D., Rao, P., Puri, K. DMatta, K.L., Surolia, A. (1992) J. Biological Chemistry, **267,** 8909- 8918.

Hammarskjold, M. L., Heimer, J., Hammarskjord, B., Sangwan, I., Albert, L., Rekosh, D. (1989) J. Virol., **63**, 1959-1966.

Hauber,J., Malir, M. H., Fenrich, R., Cullen, B. R. (1987) Proc. Natl. Acad. Sci. USA, **84**, 6364-6368.
Jakobovits, A., Smith, D. H., Jakobovits, E. B., Capon, D. J. (1988) Mol. Cell. Biol., **8**, 2555-2561.

Long, K. S., Crothers, V. (1995) Biochemistry, **34**, 8885-8895.

Loret, E. P., Vives, V., Ho, P. S., Rochat, H., Johnson, W. C. (1991) Biochemistry, **30**, 6013-6023.

Loret, E. P., P. George, W. C. Johnson, P. S. Ho (1992) Proc. Natl. Acad. Sci. USA, **89**, 9734-9738.

McClure, W. R. (1980) Proc. Natl. Acad. Sci. USA, **77**, 5634-5638

Milligan, J. F., Groebe, D. R., Witherell, G., W., Uhlenbeck, O. C. (1987) Nucleic Acids Res., **21**, 8783-8798.

Mujeeb, A. M., Bishop, K. B., Peterlin, B.M., Turck, C., Parslow, T. G., James, T. L. (1994) Proc. Natl. Acad. Sci. USA, **91**, 8248-8252.

Muesing, M. A., Smith, D. H., Capon, D. J. (1987) Cell **48**, 691-701.

Puglisi, J. D., Tan, R., Calnan, B.J., Frankel, A. D., Williamson, J.R. (1992) Science, **257**, 76-80.

Puglisi, J. D., Chen, L., Frankel, A. D., Williamson, J. R. (1993) Pro. Natl. Acad. Sci. USA, **90**, 3680-3684.

Sharp, P. A., Marciniak R. A. (1989) Cell, **59**, 229-230.

Stryer, L. (1988) Biochemistry, W. H. Freeman and Company, New York.

Tao, J., Frankel, A. D. (1992) Proc. Natl. Acad. Sci. USA , **89**, 2723-2726.

Tan, R., Frankel, A. D. (1992) Biochemistry, **31**, 10288-10294.

Weeks, K. M., Crothers, D. M. (1992) Biochemistry, **31**,10281-10287.

Weeks, K. M., Crothers, D. M. (1991) Cell, **66**, 577-588.

Weeks, C., Schultz, S. C., Steitz, T. A., Crothers, D. M. (1990) Science, **249**, 1281-1285.

Xiang, T., Goss, D. J., Diem, M. (1993). Biophysical Journal, **65**, 1255-1261.