AN ABSTRACT OF THE THESIS OF

Paul Staskus for the degree of Doctor of Philosophy in Biochemistry and Biophysics presented on January 9, 1987.

Title: A Double-Stranded Secondary Structure for Hyaluronic Acid in Aqueous-Organic Solvent as Revealed by Vacuum Ultraviolet Circular Dichroism

Abstract approved: Redacted for privacy

W. Curtis Johnson, Jr.

The association of hyaluronic acid (HA) into a double stranded structure has been investigated by circular dichroism (CD) spectroscopy into the vacuum ultraviolet region. The CD of HA changes dramatically, monitoring a cooperative transition as the dielectric constant of an aqueous solution is reduced by adding organic solvents. This structural transition results in a high intensity CD band at 190 nm, indicating an ordered structure in the mixed solvent. Heating hyaluronic acid in the mixed solvent also causes a cooperative transition reducing the CD to that found for the polymer in aqueous solution. The CD for the polymer in aqueous solution is not very different from some of the CD's for the constituent monomers, and the CD of the monomers is similar in both aqueous solution and mixed solvent. Furthermore, heating HA in aqueous solution results in small, noncooperative changes in the CD spectrum. The CD as dielectric constant is reduced exhibits isodichroic points, showing

that there are two states. Singular value decomposition of the CD as a function of organic solvent has only two principal components, confirming that there are only two states.

Oligomers of HA show a chain length effect with approximately nine disaccharides required for the structural transition as a function of organic solvent. This proves that the transition is indeed cooperative. The concentration dependence for the transition was investigated in detail for the 12 mer and 16 mer, and it was possible to prevent completely the transition in mixed solvent with sufficient dilution of the oligomer. The CD data in mixed solvent as a function of oligomer concentration were fit with various models for association of two and more strands. Simplex directed search methods were used to investigate the vector space of unknown model parameters, and two strand models were shown to consistently give a better fit. A cooperative two-strand zipper model which allows relative sliding of the chains gave the smallest error and produced the following thermodynamic parameters for the ordered structure: enthalpy of growth, -1000 ± 300 cal/mole; entropy of growth, -2.25 ± 1.30 eu/mole: enthalpy of initiation, -20,000 ± 3000 cal/mole; entropy of initiation, -71 ± 15 eu/mole. These results are consistent with a double-stranded and helical structure for hyaluronic acid in solutions of reduced dielectric constant.

A Double-Stranded Secondary Structure for Hyaluronic Acid in Aqueous-Organic Solvent as Revealed by Vacuum Ultraviolet Circular Dichroism

by

Paul Staskus

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed January 9, 1987

Commencement June 1987

APPROVED:

Redacted for privacy

Professor of Biochemistry and Biophysics in charge of major

Redacted for privacy

Chairman of Department of Biochemistry and Biophysics

Redacted for privacy

Dean of Graduate School

Date thesis is presented January 9, 1987

Typed by Barbara Hanson for Paul Staskus

ACKNOWLEDGMENTS

I wish to thank Professor W. Curtis Johnson, Jr. for his support, guidance and encouragement. Many others are deserving of recognition here, but in particular Dr. Walter A. Baase and Dr. Gary C. Causley, for their numerous helpful ideas. And to my mother and father, a special thank you for never losing faith in their children.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

A DOUBLE-STRANDED SECONDARY STRUCTURE FOR HYALURONIC ACID IN AQUEOUS-ORGANIC SOLVENT AS REVEALED BY VACUUM ULTRAVIOLET CIRCULAR DICHROISM

CHAPTER I: INTRODUCTION

The glycosaminoglycans

Hyaluronic acid (HA) is a member of the class of animal tissue polysaccharides known as the glycosaminoglycans. These polymers are named for their copolymeric structures which typically contain uronic acid residues (D-glucuronic and/or L-iduronic acid) alternating with hexosamine residues (D-glucosamine or D-galactosamine). Structural features of the glycosaminoglycans are summarized in Table I.1. Certain of the glycosaminoglycans in Table I.1 are structural isomers. For example, chondroitin and HA are identical except for the configuration at carbon 4 of their hexosamine residues, with HA containing glucosamine and chondroitin containing galactosamine. Chondroitin and dermatan are related in a similar manner. All of the acidic residues in chondroitin are glucuronic acid, while some of those in dermatan are its epimer at carbon 5, iduronic acid. Such relationships may be exploited in order to determine the importance of various structural elements in maintaining a particular physicochemical characteristic which may reflect conformational order in the polymer. HA occupies a unique position among the glycosaminoglycans in several respects. It is the longest polymer of the group, with a mass as large as several million daltons, and is one

Table I.1 Structural properties of glycosaminoglycans

Polymer	repeating residues A	B	glycosidic linkages A→B	B→A	points of sulfation
Hyaluronic acid	D-glucuronic acid	D-glucosamine	β1→3	β1→4	---
Chondroitin	D-glucuronic acid	D-galactosamine	β1→3	β1→4	galactosamine C4 galactosamine C6
Dermatan	D-glucuronic acid L-iduronic acid	D-galactosamine	β1→3 α1→3	β1→4	galactosamine C4 galactosamine C6 iduronic acid C2
Heparan	D-glucuronic acid L-iduronic acid	D-glucosamine	β1→4 α1→4	α1→4	glucosamine N glucosamine C6 iduronic acid C2
Keratan	D-galactose	D-glucosamine	β1→4	α1→3	glucosamine C6

of few polymers attaining such size which is not synthesized from a template (Laurent, 1966). It is the only one of the more common glycosaminoglycans not yet demonstrated to be covalently attached to protein in its natural form (Mason et al., 1982). It is also the only glycosaminoglycan synthesized not only by eukaryotic cells, but also by certain bacterial species (Sugahara et al., 1979).

The glycosaminoglycans are widely distributed throughout the animal kingdom, where they are perceived to perform certain common functions related to their physicochemical properties. These include maintenance of connective tissue integrity, control of macromolecular distribution by steric exclusion, and water and ion binding (Mathews, 1975). Such functions do not warrant the variability observed among the glycosaminoglycans in structure or tissue distribution, and it is likely that the glycosaminoglycans also interact specifically with other cell and tissue components (for a review, see Lindahl and Hook, 1978).

The role of heparin in the control of blood clotting is a good example of this. Heparin is structurally the most complex glycosaminoglycan, possessing both types of uronic acid and considerable microheterogeneity in sulfation of its residues (summarized in Table I.1). Part of the heparin structure has a high affinity for antithrombin and can affect its binding to and inactivation of thrombin, an important protease in the blood clotting cascade (Nordenman et al., 1978; Lindahl et al., 1979). Curiously, the primary structure of heparin is not precisely defined, but is created from a regular precursor polymer through a series of enzymatic modifications involving an ordered reaction sequence which does not

necessarily proceed to completion for an entire polymer molecule (Hook et al., 1975; Lindahl et al., 1977). The resulting mixture of precise and random events creates the unique sequences in heparin capable of activating the protease inhibitor.

Several physiological functions have been proposed for HA, such as that of a shock absorber and lubricant in articular cartilage (Preston et al., 1965), a filter regulating flow in the vitreous of the eye (Balazs and Gibbs, 1970), and a modifier of cell behavior during differentiation (Derby, 1978) and tumor invasion (Toole et al., 1979). The molecule has also been suggested to serve as a mechano-electrical transducer in the inner ear (Barrett, 1975, 1976; Barrett and Harrington, 1977), an effector of tissue deformability during pregnancy (Golichowski et al., 1980), a maintainer of tissue spaces during development for use as passageways during cellular migrations (Fisher and Solursh, 1977), an enhancer of diffusion of small molecules (Napier and Hadler, 1978), and an organizer of proteoglycan molecules in cartilage (Hascall, 1977). This wealth of potential functions may be regarded as due in part to the simplicity of HA's primary structure. In fact, it is the simplest in structure of all glycosaminoglycans, being depicted in Figure I.1 as a strictly alternating copolymer of 2-N-acetyl-2-deoxy-D-glucopyranose (N-acetyl-D-glucosamine, GlcNAc) residues in β-1,4 linkage with D-glucopyranoside uronic acid (D-glucuronic acid, GlcUA) residues in β-1,3 linkage (Meyer, 1958). It does not contain any sulfate groups which create microheterogeneity and increase linear charge density in other glycosaminoglycans. Ironically, the relative structural simplicity and low information content of HA and other

Figure I.1 HA primary structure. A pair of disaccharides in the repeating structure of HA. The molecule is poly[(1->3)-O-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-(1->4)-O-β-D-glucopyranuronosyl]. Shown as dotted lines are the four hydrogen bonds proposed to account for the chain stiffness of HA in aqueous solution (Atkins et al., 1980). N(1), N-acetylglucosamine residue at reducing end. G, glucuronic acid residue (Redrawn from Scott et al., 1984).

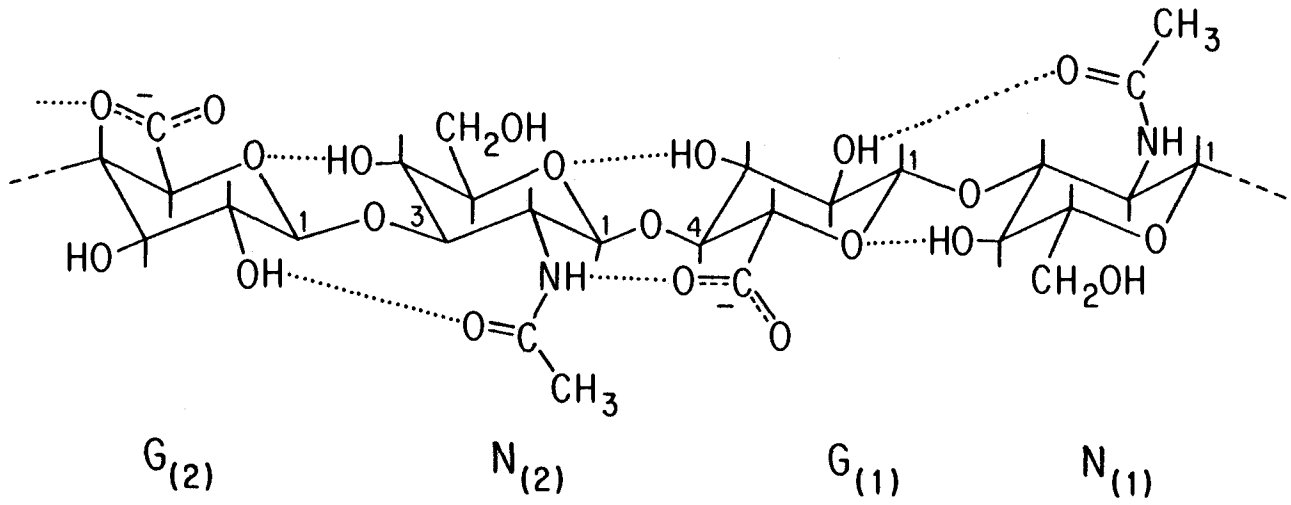

Figure I.1

glycosaminoglycans relative to proteins and nucleic acids has likely slowed research concerning their pivotal roles in such complex processes as embryological development and tissue organization.

The binding of proteoglycan core protein by HA, an important function in connective tissue organization, demonstrates that polysaccharide structural complexity is not prerequisite to specific interaction of a glycosaminoglycan with a protein. Cartilage proteoglycan, itself a large molecule containing protein with covalently attached glycosaminoglycan chains of chondroitin sulfate and keratan sulfate, binds HA in a manner which resists competition by the isomeric glycosaminoglycan chondroitin, as well as other polyanions such as dextran sulfate, alginate and DNA (Hardingham and Muir, 1972). Using exclusion chromatography and viscometry to monitor the size of the complex formed by bovine nasal cartilage proteoglycan with either HA or its shorter segments, Hascall and Heinegard (1974) have shown that oligomers of HA that are 10 saccharides and longer can compete with the polymer for proteoglycan binding, whereas shorter oligomers are ineffective.

Conformational order in HA

Most of the functions proposed for HA have some basis in the expansive nature of the molecule, which can occupy a domain 10^3-10^4 times the volume of the polymer chain itself (Balazs, 1958). The polymer has been characterized in solution as a random coil with some stiffness (Laurent, 1970; Preston et al., 1965). Studies of the binding of HA to proteoglycan (Faltz et al., 1979; McMurtrey et al.,

1979; Cleland, 1979; Christner et al., 1979) or of its interaction with and modification of the behavior of cells (Turley et al., 1985; Goldberg and Toole, 1984; Underhill, 1982; Wasteson et al., 1973; Turley and Roth, 1979), have not resulted in any proposal of an ordered structure in HA being necessary for its function. However, considerable evidence of conformational order in HA solutions has accumulated. This includes the results of studies based on viscometry (Swann and Caulfield, 1975), nuclear magnetic resonance (NMR) (Darke et al., 1975), periodate oxidation kinetics (Scott and Tigwell, 1978), rheology (Morris et al., 1980; Welsh et al., 1980) and optical activity (Chakrabarti and Balazs, 1973; Cowman et al., 1981, 1983; Park and Chakrabarti, 1978).

Scott and Tigwell (1978) have investigated the resistance of HA and chondroitin to oxidation by periodate ion. The uronic acid residues in glycosaminoglycans contain a glycol functional group at carbons 2 and 3 of the hexose ring which potentially will react with periodate. In glucuronic acid, the hexose ring is likely in the C1 configuration, which places the carboxyl group at carbon 5 in an equatorial position. The hydroxyl groups at carbons 2 and 3 are also equatorial, and thus favorably disposed for reaction with periodate ion. Iduronic acid residues may adopt the 1C ring configuration in order to maintain the carboxyl group in an equatorial position. This would place the hydroxyl groups at carbons 2 and 3 on opposite sides of the ring in axial positions, and should make periodate oxidation of the uronic acid more difficult. It is therefore anticipated that glycosaminoglycans containing glucuronic acid, such as hyaluronate and chondroitin, may be oxidized by periodate more quickly than those

glycosaminoglycans containing some iduronic acid, such as dermatan and heparan. The opposite is found to be the case (Scott, 1968; Fransson, 1974). The resistance of HA and the chondroitin sulfates to periodate oxidation, when compared to various polyuronic acids, dermatan sulfate and heparan sulfate, was explained by proposing a hydrogen bond network along the polymer chain involving the glycol groups in HA and chondroitin sulfate which is not possible in heparan or dermatan. The bonds in this scheme are depicted in Figure I.1. The hydroxyl group on carbon 2 of glucuronic acid is bonded to the acetamido oxygen of the amino sugar, stabilizing the glucuronidic linkage conformation in the HA backbone. Similarly, the hydroxyl group on carbon 3 is bonded to the amino sugar ring oxygen, stabilizing the hexosaminidic linkage in the backbone. Scott and co-workers have conducted NMR measurements in dimethyl sulfoxide solution of glycosaminoglycans and their oligomers (Heatley et al., 1982; Scott et al., 1984, 1981; Scott and Heatley, 1982). The proton chemical shift patterns obtained support the existence of a hydrogen bonding scheme such as that depicted in Figure I.1 for hyaluronic acid and the chondroitin sulfates, but not dermatan sulfate, heparan sulfate or keratan sulfate. Such evidence has not been confirmed in aqueous solution (Cowman et al., 1984).

Hydrogen bonding within the backbone of HA, such as that which can explain the reduced rate of oxidation by periodate, has also been proposed in model building studies of the polymer performed to explain the origins of x-ray fiber diffraction patterns. Modelling studies of HA indicated that most of its conformational space is inaccessible due to steric hindrance (Rees, 1969). However, the remaining conformational space contains many local energy minima, and rather

large variations in helical structure are possible (Potenzone and Hopfinger, 1975). Indeed, HA has an unusual extent of variability in conformation and packing of strands in the solid state, as functions of temperature, relative humidity, pH and counterion (Sheehan et al., 1975; Guss et al., 1975; Winter et al., 1975; Winter and Arnott, 1977; Arnott et al., 1983; Mitra et al., 1983b,c). The fiber conformations proposed for HA and the conditions of their formation are presented in Table I.2.

There are three or four major diffraction patterns for HA fibers. The conformational models corresponding to them include a left-handed three-fold helix (three disaccharides per turn) found in the presence of divalent cations such as Ca^{++} or Sr^{++} (Winter and Arnott, 1977), a two-fold helix (two disaccharides per turn) proposed to exist under condition of low water content in the presence of acid (Atkins et al., 1980), and left-handed four-fold helices (four disaccharides per turn) formed in the presence of monovalent counterions such as Na^{+} and K^{+} (Sheehan et al., 1977; Mitra et al., 1983b,c). The three-fold helix is rather extended, with a pitch of .95 nm per disaccharide. Similar conformations are found in other glycosaminoglycans in the presence of monovalent ions, including chondroitin sulfate (Cael et al., 1978; Atkins et al., 1974a) and dermatan sulfate (Mitra et al., 1983a). The two-fold helix of HA found in the presence of acid is also quite extended with a pitch of .98 nm per disaccharide. Such a conformation is observed in chondroitin sulfate in the presence of Ca^{++} ion (Cael et al., 1978). The left-handed four-fold helices proposed to exist in HA fibers formed in the presence of monovalent cations are not found in any other glycosaminoglycans. One of these helices has a more

Table I.2 Conformational parameters of hyaluronic acid fiber helices[a]

helix symmetry	disaccharide pitch (nm)	counterion	hydrogen bonds[b]	References
2_1	0.98	free acid	05(A)-04(B) 02(A)-OAc(B) 06(A)-N(B) 03(A)-05(B)	Atkins et al., 1972, 1980
3_2	0.95	Ca^{++}, Sr^{++}	03(A)-05(B) 05(A)-04(B)	Winter et al., 1975 Sheehan et al., 1975 Winter and Arnott, 1977
4_3	0.85, 0.89	Na^+, K^+	05(A)-04(B) 06(A)-N(B)	Guss et al., 1975 Dea et al., 1973 Atkins et al., 1973
4_3	0.95	K^+	05(A)-04(B)	Mitra et al., 1983
4_3	0.82	K^+, NH_4^+, Rb^+, Cs^+ at low pH	06(A)-N(B) 02(A)-OAc(B) 05(A)-04(B) 06(A)-06(A)[c]	Sheehan et al., 1977 Arnott et al., 1983

[a] HA typically packs as single helices with varying degrees of contact with several neighboring chains. The one exception here is the 4_3 structure with .82 nm pitch. Its chains pack as antiparallel coaxial double helices, with only one of the seven hydrogen bonding contacts per disaccharide not being confined to the double helix.

[b] intrachain hydrogen bonds between glucuronic acid (A) and N-acetyl glucosamine (B) residues, restricting glycosidic linkage flexibility.

[c] interchain, intra double helix bond.

contracted structure with a pitch of only .82 nm per disaccharide. Its corresponding fiber was formed in the presence of K^+, NH_4^+, Rb^+ or Cs^+ at low pH (Sheehan et al., 1977). Similarly contracted structural models accounting for x-ray patterns from fibers of plant polysaccharides such as the carrageenans are double-helical, with two polymer strands winding about a common axis (Arnott et al., 1974b; Anderson et al., 1969). The four-fold HA structure may be included in this category, with placement of the carboxyl groups of uronic acid residues toward the center of the double helix accounting for the requirement of low pH. A double-helical structure for HA had been considered previously to account for a different diffraction pattern (Dea et al., 1973). Subsequent structure factor calculations had shown a four-fold single helix model superior in accounting for the data (Guss et al., 1975). The double helix proposed by Sheehan et al. (1977) is itself controversial in that the structure contains no apparent means of holding the HA chains together, and none of the intramolecular hydrogen bonds across the backbone glycosidic linkages commonly seen in other HA structures are present. A recent reanalysis of these diffraction data by Arnott et al. (1983), allowing for modest changes in pyranose ring torsion angles and endocyclic bond angles resulted in a new conformation which is still double-helical. The new conformation contains seven hydrogen bonds per disaccharide, including three of the four intrachain bonds across glycosidic linkages depicted in Figure I.1. The model also proposes a direct hydrogen bond linking uronic acid carboxyl groups between the paired strands of HA within the double helix, as well as a water bridge.

Hydrogen bonding across the glycosidic linkages in HA such as that

depicted in Figure I.1 can be anticipated to provide a stiffening influence in the polymer. This stiffness would help account for the extended structure of the molecule reflected in its solution viscosity. HA can contribute over 80% of the viscosity of biological fluids at weight concentrations less than 0.1% (Balazs and Gibbs, 1970). Segmental stiffness should also reduce the entropic cost of chain-chain association with dynamic network formation. Such a transient tertiary or higher order structure is suggested by the viscoelastic properties of HA solutions, in which intermolecule coupling begins at an unusually low degree of coil overlap (Morris et al., 1980b). This network character is accentuated by reduction of pH or water content, with formation of a viscoelastic putty at higher concentrations of polymer in aqueous solution at pH 2.5 (Balazs, 1966).

Rheological studies have demonstrated that network formation in HA solutions is specifically inhibited by oligomers of the same copolymeric sequence (Welsh et al., 1980). The smallest oligomers tested, being 2-4 disaccharides long, had no effect on network formation. Material averaging 60 disaccharides in length inhibited or abolished HA polymer network formation under the experimental conditions used. Larger segments of HA, on the order of 400 disaccharides long, actually increased the network character of polymer solutions to which they were added. The interpretation proposed for these results was analogous to that of other polysaccharide systems such as agarose, alginate and carrageenan (Rees and Welsh, 1977), in which formation of local double-helical junction zones may be accompanied by higher order association to effectively

cross-link polymer molecules. Presumably, HA oligomers of 60 disaccharides are long enough to participate in junction zone formation, but not long enough to be involved in more than one such zone at a time, thus inhibiting network organization (Morris et al., 1980b).

Results of chiroptical studies also suggest conformational order in HA solutions. Stone pioneered this approach by studying optical rotatory dispersion induced in symmetric dye molecule transitions upon binding to various glycosaminoglycans (Stone, 1964, 1965). Later circular dichroism (CD) measurements revealed a correlation between the sign of the CD in the energy region common to π-π^* transitions and amino sugar linkage. Amino sugars linked 4-1 in glycosaminoglycans have a positive CD band near 190 nm, while those linked at carbon three have a negative CD band centered at a wavelength shorter than 185 nm (Stone, 1969, 1971). Stone suggested that the intensity of the π-π^* band relative to an n-π^* band at longer wavelength reflected the degree of order in a glycosaminoglycan.

The CD spectrum of HA polymer has been shown to differ from either a simple algebraic sum of its monomer spectra (Buffington et al., 1977), or from spectra of its tetra- or hexasaccharides (Chakrabarti and Balazs, 1973). Cowman et al. (1981, 1983) have demonstrated the enhanced CD of HA polymer in aqueous solution to arise from a local effect. Oligomers of HA have a CD in the acetamido n-π^* transition region which reflects the ratio of hexosaminidic linkages and total hexosamine residues in them. The n-π^* CD is therefore describable as an end effect. The CD in the π-π^* transition region of these same oligomers can be described as an algebraic sum of contributions from

hexosaminidic and glucuronidic linkages. This is apparent even in the shortest oligomers of HA, and so does not suggest any long range conformational order in the polymer which could play a role in its function.

Park and Chakrabarti (1977, 1978) observed a transition in the near ultraviolet CD spectrum of HA upon lowering the pH and raising the fraction of organic component in a mixed solvent sample solution. They showed the transition to be reversed upon raising the pH or temperature of the sample, and cooperative in nature. The transition was not observed either in chemically degraded HA or in its isomer, chondroitin. Furthermore, the chiroptical transition could be correlated with viscosity changes in the HA solution (Park and Chakrabarti, 1978b). The CD spectrum of HA in aqueous-organic solution at low pH resembles the spectrum of a HA film (Buffington et al., 1977). It was therefore suggested to represent HA in one of the four-fold helical conformations from x-ray diffraction studies.

Lower pH and decreased dielectric constant of the solution are two conditions that favor network formation in HA (Morris et al., 1980b). This suggests that the CD transition of HA in a mixed solvent system reflects chain association, with formation of a structure such as that predicted from analysis of x-ray diffraction patterns obtained from fibers (Arnott et al., 1983). To investigate this possibility, we have extended the measurements of HA in aqueous-organic solvent into the vacuum ultraviolet region, where the CD signal changes are more pronounced. We have isolated oligomers of HA and performed CD measurements to clarify the nature of the chiroptical transition displayed by HA in acidic aqueous-organic solvents. We are able to

show that spectra describing the transition contain only two components of information above the noise. This tells us that the chromophores of HA responsible for the CD can occupy two different environments as the molecule's secondary structure changes. We will therefore fit our transition data using models that allow HA residues to exist in two states. The transition is clearly demonstrated by HA oligomers 9 or more disaccharides long, but not by shorter oligomers. This dependence on chain length is therefore distinct from the end effect in HA CD studied by Cowman et al. (1981, 1983) in aqueous solution. It indicates that many residues of HA are cooperating to form an ordered structure in aqueous-organic solvent, and defines a critical length in the HA molecule necessary for the transition to take place. We will investigate the quality of fit which results when we employ cooperative models with two states for polymer residues to describe our data, as opposed to simpler "all-or-none" transition models. We will also investigate models involving multi-stranded structures, as we observe the transition to be sensitive to the concentrations of the oligomer samples, suggesting chain association. Chapter IV presents our work on oligomers of HA and a data analysis approach for determining thermodynamic and other parameters of the transition described here. It follows a more detailed description of the transition for the "infinite" polymer in Chapter II, and a discussion of transition cooperativity, which is necessary to adequately describe the data for the oligomers, in Chapter III.

CHAPTER II: THE CHIROPTICAL TRANSITION OF THE HYALURONIC ACID POLYMER

INTRODUCTION

Chiroptical studies of glycosaminoglycans (GAG's) were pioneered by Stone in the 1960's, using cationic dyes which aggregate on the polyanions (1964, 1965, 1967). Wavelength and magnitude perturbations of dye absorption bands and induced optical rotatory dispersion are observed when the dye molecules are complexed with the polysaccharides. Such effects had previously been observed in dyes bound to polypeptides (Blout and Stryer, 1959) and polynucleotides (Neville and Bradley, 1961). Since the effects in those systems were dependent upon a helical conformation of the polymer, Stone's results suggested that the GAG's could also exist in helical forms, at least when complexed with the dye molecules. The studies of GAG-dye complexes were extended to direct measurement of Cotton effects in GAG's at frequencies typical of amide π-π^* and n-π^* transitions (Stone, 1969, 1971). Certain optical properties shared by the GAG's and amino sugar derivatives appeared to derive from the amide chromophore. Among the GAG's, gross features of CD spectra were correlated with the types of backbone linkages to the amino sugars.

CD measurements of GAG's were first extended into the vacuum ultraviolet region by Buffington et al. (1977) with their spectrum of HA. Their results verified nonadditivity of the polymer spectrum relative to the monomer spectra. Enhanced CD of the HA polymer relative to its tetra- and hexasaccharide digestion products had previously been noted by Chakrabarti and Balazs (1973). They

suggested that the difference was due to adoption of an ordered conformation by the polymer. Buffington et al. (1977) compared changes in the CD of HA as a function of pH with those of the monomer glucuronic acid. In contrast, they proposed that the effect of pH on the polymer spectrum could be accounted for by changes in the monomer transitions resulting directly from protonation of the carboxylate groups, without invoking an ordered polymer secondary structure.

Through studies of sodium hyaluronate (NaHA) oligomers, Cowman et al. (1981) have shown that the enhanced CD of NaHA polymer relative to the monosaccharides in the n-π^* transition region can be expressed using an end effect model, in which the enhanced CD reflects the content of β-1,4 linkages in a given molecule. Study of the vacuum ultraviolet CD bands (Cowman et al., 1983) showed that the spectrum of a long oligomer of NaHA could be described as the average of two spectra, each corresponding to end residue contributions in the two different types of alternating disaccharide oligomers, oligo(NaGlcUA-GlcNAc) and oligo(GlcNAc-NaGlcUA). No evidence for a polymer conformation requiring cooperative interaction among residues was observed.

Park and Chakrabarti (1977, 1978a,b) observed a dramatic effect in the CD spectrum of HA upon reducing the dielectric constant of the bulk medium by introducing various organic solvents, and increasing the proton concentration to millimolar quantities. The resultant spectrum is reminiscent of that for NaHA film (Buffington et al., 1977), in which the negative CD band typical of GAG's in the n-π^* transition region near 210 nm is no longer resolved, but appears as a shoulder of an intense negative band at shorter wavelengths.

Coincidence of an inflection in the absorption spectrum of the NaHA film with the intense negative CD confirmed its assignment to a π-π^* transition, presumably of the acetamido chromophore (Buffington et al., 1977). The large rotational strength of the chromophore was ascribed to decreased rotational flexibility of the acetamido group as it participates in hydrogen bonding with the carboxyl group. The proposed structure was that of a four-fold helix deduced from x-ray diffraction studies of similar NaHA films (Guss et al., 1975).

We have performed CD measurements in the vacuum ultraviolet for HA in aqueous solution and mixed aqueous-organic solvent at low pH. Our spectra clearly show the intense negative CD of HA in the mixed solvent to be centered at 188 nm. Our absorption data indicate that this band is due to a π-π^* transition in the acetamido group of HA. We can observe the effects of solvent composition, temperature, and pH on the CD of HA in the π-π^* transition region. Our results are generally in agreement with those of Park and Chakrabarti (1978a,b) obtained in the longer wavelength region of the HA spectrum. However, there are important differences between the two sets of data which are likely due to experimental factors in the Park and Chakrabarti measurements such as sample aggregation and instrumental limitations.

We can take spectra of HA recorded as a function of solvent composition at low pH, and recombine them in a linear fashion to create a basis set whose elements are mutually perpendicular in a vectorial sense (i.e. the dot product of any two is the zero vector). The resulting basis set, obtained by singular value decomposition (Forsythe et al., 1977) of the matrix formed from the data spectra,

has only two significant elements. This demonstrates that the chromophores of HA responsible for the changes in CD observed as a function of solvent composition have only two allowable environments. We therefore have utilized two-state transition models to fit data which describe melting of the mixed solvent HA structure reflected in the CD.

A suitable description of the HA CD spectra recorded during melting requires three significant elements of the mutually perpendicular basis set formed from them. We can account for the necessity of the third component by considering the effect of temperature on the CD spectrum of HA recorded in aqueous solution at low pH. However, we find that two-state models do not adequately describe the melting data, even when they are corrected for the above temperature effect. This failure may be due to temperature dependence of the transition enthalpy, or the involvement of more than one energy in the transition process. Two-state models allow for only one enthalpy and entropy in describing the transition as an "all-or-none" process. Since we are dealing with a polymer molecule, we must recognize the potential for separate initiation and growth phases with distinct energies in any structural transition. We can verify the interaction between HA residues during the transition in mixed solvent, by a systematic investigation of HA oligomers.

MATERIALS AND METHODS

Monovalent salts of HA polymer were purchased from Sigma Chemical Co. (bovine vitreous humor, grade IV) and Calbiochem-Behring. Material molecular weights were estimated from intrinsic viscosity measurements in 0.2 M NaCl solution at 25 degrees C, using the relation

$$[\eta] = 0.228\ MW^{0.816} \qquad (II.1)$$

(Cleland and Wang, 1970). Viscosity measurements were performed in a semimicro version of the floating rotor couette viscometer of Zimm and Crothers (1962). Details of viscometer operation are in Appendix VI. Molecular weight of the Calbiochem HA was determined to be 170,000 daltons, and that of the Sigma material was 240,000.

Unless stated otherwise, mixed solvent refers to a solution containing ethanol at a concentration of 45% by volume to volume of solution, or 7.92 M. Acidic mixed solvent has an aqueous component with a pH of 2.5, containing phosphoric acid. HA polymer was prepared for spectroscopy by deionizing an aqueous solution with a strong cation exchange resin in the proton form (AG 50W, Bio-Rad), followed by lyophilization and solvation in water. Solutions of polymer sample were typically unbuffered, using phosphoric acid to lower pH and containing one of several different nonpolar solvents at various concentrations. Nonpolar solvent was gravimetrically added to an acidic aqueous solution of HA, with further addition of water to known final volume after mixing. Alternatively, aqueous solutions of

polymer were dialyzed four times against a 500-fold excess of the desired solvent. All water used in these experiments was glass distilled.

Circular dichroism spectra were recorded using a vacuum ultraviolet spectrograph described previously (Johnson, 1971). The CD signal was calibrated with a known concentration of (+)-10-camphorsulfonic acid (Aldrich) in a 1 mm cell, assuming a $\Delta\varepsilon$ of 2.41 $M^{-1}cm^{-1}$ at 290.5 nm. Wavelength calibration of the spectrograph was accomplished by using a mercury arc lamp as the light source, and comparing the monochromator wavelength position with mercury emission lines at 185.0, 253.7, 296.7, and 365.0 nm. Many of the data were collected by a single board microcomputer (SUPERKIM, Lamar Instruments, Redondo Beach, CA) assembled in this laboratory and equipped with a 12 bit analog to digital converter (Analog Devices AD 574 KD). The computer was programmed in machine code to operate the converter at approximately 2 KHz, divided between two signal inputs. More detail of computer assembly and programming can be found in Appendix V. Most commonly, a spectral bandwidth of 1.6 nm and scan rate of 0.5 nm per minute were used, with the signal being averaged over a period of one minute per data point. Samples were contained in cylindrical cells (Hellma) with nominal pathlengths of 0.01, 0.05, 0.1, 0.2, and 1.0 mm. For all CD measurements, total optical density (OD) of sample, solvent, and cell was less than 1.0. Pathlengths of cells 0.1 mm and shorter were determined interferometrically (Bree and Lyons, 1956). Longer pathlengths were checked by absorption measurements of a sample of known concentration and extinction coefficient. Spectra recorded using a thin cell apparatus, in which

the sample is sandwiched between two quartz windows with an effective pathlength of 2-20 microns, were scaled to provide agreement with results from longer pathlength cells in the spectral region common to the two measurements. Flow linear dichroism (LD) measurements were performed using the apparatus described by Causley and Johnson (1982). The pathlength of the assembled flow cell was 35 microns.

HA concentrations were based on the colorimetric determination of glucuronic acid (Bitter and Muir, 1962). For spectroscopic assay of concentration, absorption measurements were performed using a Cary 15 or Cary 219 spectrometer with nitrogen purge. The results were compared with colorimetric determinations of HA concentration to establish extinction coefficients for the polymer at 190 nm under various solution conditions.

Sample temperature was controlled by cell placement in a brass jacket through which glycol-water of constant temperature was cycled by a Lauda recirculating bath. Thermal contact between cell and jacket was maintained under vacuum by Thermocote paste (Thermalloy, Dallas, TX). The paste was also used to contact thermistor probes (Yellow Springs Instruments) with both the brass jacket and the edge of the cylindrical cell face, for temperature monitoring. A third probe was temporarily attached to the center of the cell face, to establish the relationship between the temperatures at the edge and the center. This relation was linear, and insensitive to cell pathlength or sample volume within the cell. Probe resistances were calibrated with a thermometer in the temperature range from 0 to 50 degrees C, and the resistance was then measured directly.

Series of CD spectra recorded as a function of temperature or

solvent composition were analyzed for orthogonal components (Hennessey and Johnson, 1981) using a matrix decomposition technique called singular value decomposition (Forsythe et al., 1977). Significance of the basis spectra so created was judged by visual inspection of the bases and their associated eigenvalues. Approximations of the original data spectra were then created using only the significant components of the basis set. These reconstructed spectra were observed to fit the original data within the noise level of the CD measurements.

RESULTS AND DISCUSSION

Monomer Spectra

CD spectra of the HA monomers are presented in figure II.1. The monomer spectra contain information concerning the chromophores present in HA and are worthy of study in their own right. However, we present them here primarily as a reference point for comparison with the polymer spectra. Those spectral features of polymeric HA not found in the monomer spectra may be important in reflecting conformational order and the existence of chromophore environments not possible in the monomers. Care must be exercised in interpreting such differences since they may arise from other sources as well, such as the formation of new covalent features (glycosidic bonds) and chromophores (acetal from hemiacetal) which should influence the interactions among parts of the molecule which produce optical activity. We will see that there are features of HA polymer CD spectra that are unexpected from consideration of the monomer CD. More importantly, the changes observed in polymer spectra upon altering the dielectric constant of the solvent are far more dramatic than the corresponding changes in the monomer spectra.

Figure II.1a presents CD spectra of glucuronic acid (GlcUA) under the solvent conditions used in this study. The spectrum of GlcUA measured in aqueous solution at pH 2.5 (solid curve) is in agreement at wavelengths longer than 190 nm with the result of Buffington et al. (1977). In addition, we have extended our solution spectrum further into the ultraviolet, and observe a negative band centered at 183 nm

Figure II.1 CD spectra of HA monomers.

a) CD spectrum of GlcUA recorded in an aqueous solution 12.5 mM in NaH_2PO_4, 7.5 mM in H_3PO_4, pH 2.5 in the absence (———) and in the presence (— — —) of 45% v/v ethanol (7.92 M). Spectra were recorded at 24 degrees C in 52.4 and 11.9μ cylindrical cells at 51 mM concentration of GlcUA.

b) CD spectrum of GlcNAc in the same solution condition as above, in the absence (———) and in the presence (— — —) of 45% v/v ethanol. GlcNAc concentration was approximately 15 mM.

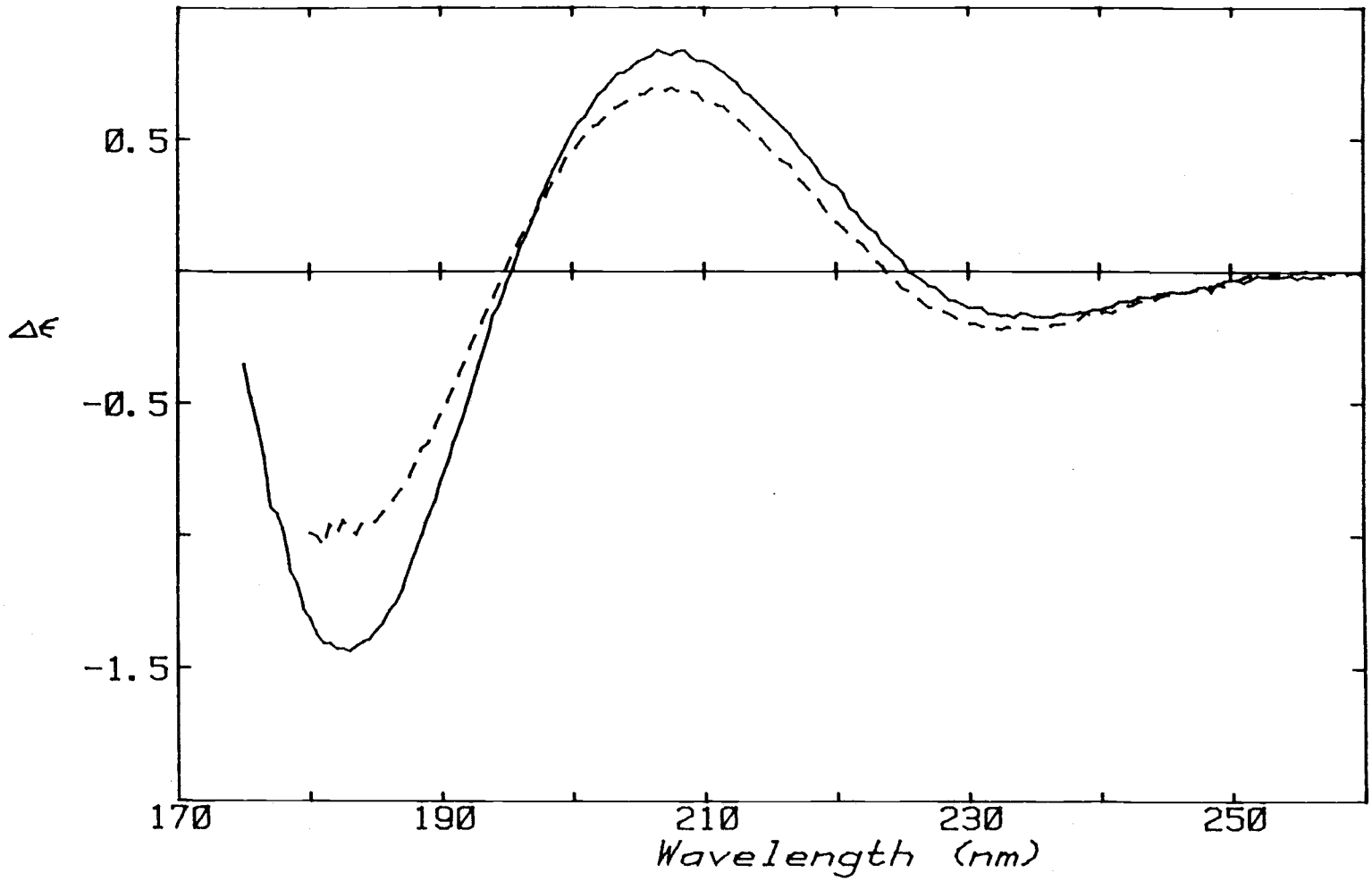

Figure II.1a

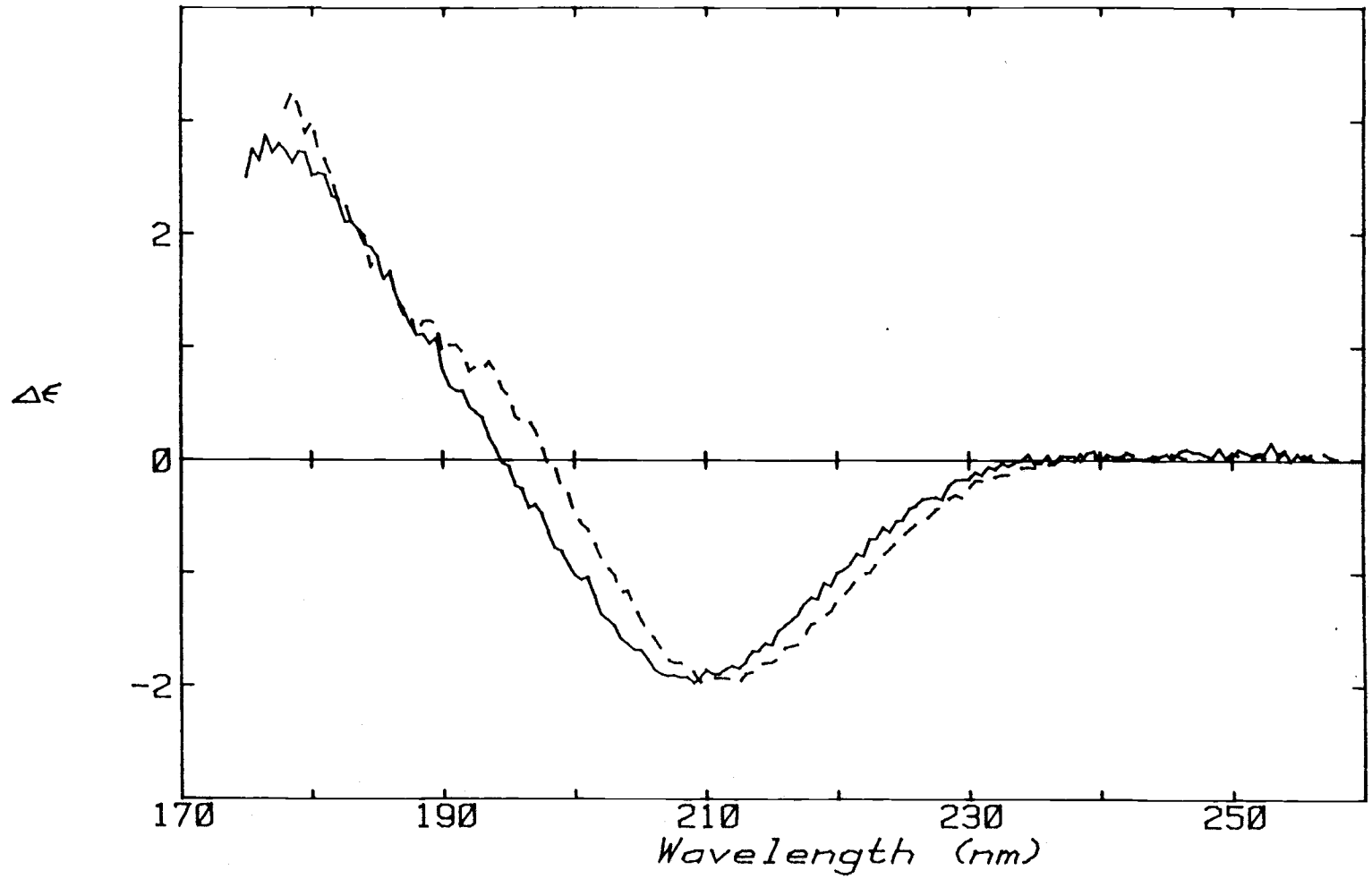

Figure II.1b

with $\Delta\varepsilon$ of -1.42, corresponding in position and sign to the band observed by Buffington et al. (1977) in GlcUA film. The longer wavelength bands of opposing sign are believed to arise from n-π^* transitions in different rotational isomers of the uronic acid (Listowsky et al., 1969). The 183 nm band has been interpreted as due to a carboxyl π-π^* transition in the molecule (Buffington et al., 1977). Introduction of ethanol (dashed curve, 45% v/v) induces a slight blue shift of the CD crossover at 225 nm, and a modest decrease in CD throughout the spectrum, with the effect being somewhat larger for the negative 183 nm band. There are no sign inversions or gross magnitude changes in CD bands. A change in temperature from 24 to 50 degrees C also has little effect on the GlcUA spectrum in aqueous or mixed solvent (data not shown).

Normal absorption spectra of GlcUA under these solution conditions contain little information in the same range. There is no discernible band maximum in the π-π^* transition region, although the absorption is observed to increase with an increase in pH. There is a long weak tail with an ε_{220} of approximately 65 $\underline{M}^{-1}cm^{-1}$ which extends to 230 nm in the GlcUA absorption spectrum at pH 2.5. This tail either disappears or is blue shifted approximately 10 nm in the spectrum recorded at pH 6.8 (figure II.2a).

CD spectra of 2-acetamido-2-deoxy-D-glucopyranose (GlcNAc) in acidic aqueous and acidic mixed solvents are presented in Figure II.1b. The negative CD band at 210 nm has been assigned to an n-π^* transition in the acetamido chromophore (Buffington et al., 1977). It is a common feature in the CD spectra of glycosaminoglycans, as originally reported by Stone (1969). The aqueous solution spectrum of

Figure II.2 OD spectra of HA monomers.

a) OD spectrum of GlcUA recorded in an aqueous solution containing 20 mM sodium phosphate at pH 2.5 (———) and pH 6.8 (– – –). Spectra were recorded at 24 degrees C in 1 cm and 52.4 μ cells at 2.5 and 51 mM concentration of GlcUA. Presence of 45% v/v ethanol resulted in essentially the same spectrum.

b) OD spectrum of GlcNAc in the same solution condition as above, pH 2.5 (———) and 6.8 (– – –). GlcNAc concentration was approximately 15 mM. Presence of 45% v/v ethanol resulted in the same spectrum.

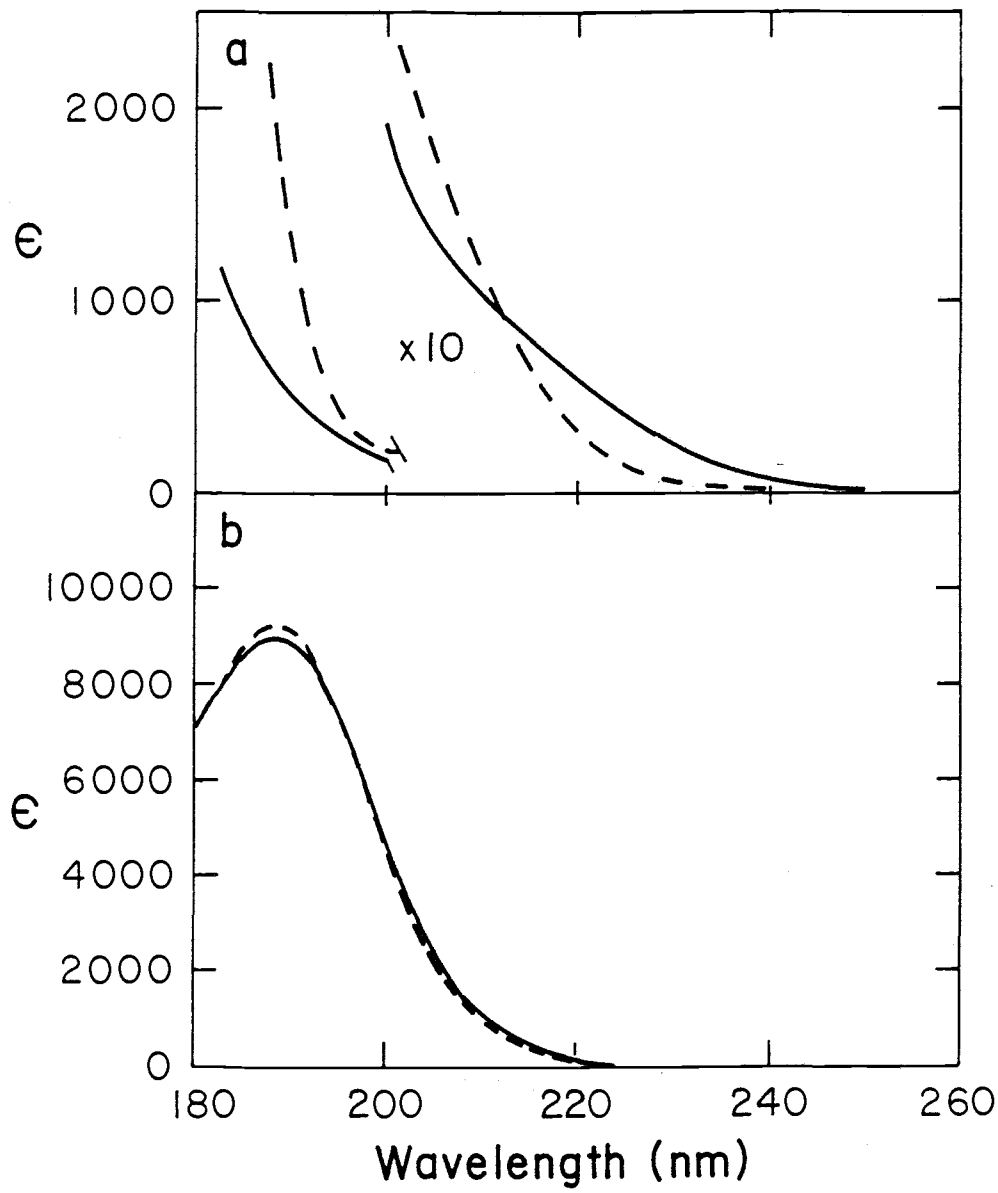
a
2000
ε
1000
0
x10
b
10000
8000
6000
ε
4000
2000
0
180
200
220
240
260
Wavelength (nm)

Figure II.2

GlcNAc (solid line in the figure) displays no clear band maximum in the π-π^* transition region. Addition of ethanol to the solution does result in the appearance of a positive shoulder in this region, however (dashed line). It coincides with a maximum in the normal absorption spectrum of the molecule at 188 nm (ε_{188} ~9000 $\underline{M}^{-1}cm^{-1}$), which is insensitive to changes in solvent composition and pH (figure II.2b). The strong absorption is consistent with the assignment of CD in this region to a π-π^* transition, by Buffington et al. (1977).

Our measurement of $\Delta\varepsilon_{210}$ is somewhat larger than that of Buffington et al. (1977) and Cowman et al. (1983), being -1.93 as opposed to -1.36. The value was determined using the dry weight of GlcNAc in the sample, which also resulted in an ε_{188} of 9000 $\underline{M}^{-1}cm^{-1}$ for the normal absorption spectrum. Using 8000 $\underline{M}^{-1}cm^{-1}$ for ε_{188} (Cowman et al., 1983) reduces $\Delta\varepsilon_{210}$ approximately 10%, but our value of 9000 $\underline{M}^{-1}cm^{-1}$ for ε_{188} is more consistent with the following data.

In examining all of our spectra of HA polymer and oligomers in acidic aqueous or mixed solvent, we find an extinction coefficient for the normal absorption of the disaccharide at 188 nm of 9500 $\underline{M}^{-1}cm^{-1}$ to be universally consistent in relating absorption to concentration determined colorimetrically (Bitter and Muir, 1962). This value is approximately the sum of the extinction coefficients for the monomers at 188 nm under these solvent conditions (i.e. 9000 $\underline{M}^{-1}cm^{-1}$ for GlcNAc and ~700 $\underline{M}^{-1}cm^{-1}$ for GlcUA). Cowman et al. (1983) determined the ε_{188} for NaHA segments in neutral aqueous solution as 11,400 $\underline{M}^{-1}cm^{-1}$. The difference appears to be due to a change in GlcUA absorption with pH, as we estimate its ε_{188} at pH 6.8 in aqueous solution to be ~2000 $\underline{M}^{-1}cm^{-1}$.

As shown in Figure II.1b, addition of ethanol to GlcNAc solutions results in a red shift of the negative CD band in the n-π^* region, with appearance of a positive CD shoulder near 190 nm. A red shift with reduction of solvent dielectric constant is common for n-π^* transitions. It has been explained (Lantzke, 1973) in terms of changes in polarity of the chromophore with the transition, and its interaction with the solvent molecules. Usually the chromophore is less polar in the excited state than it is in the ground state, since the localized lone pair electrons move to a more diffuse orbital. In a polar solvent like water, a cage of polarized solvent molecules will be set up around the ground state and the excited state will tend to induce strain in the solvent cage. Reducing the dielectric constant of the solvent decreases the degree of solvent orientation around the ground state chromophore, raising its energy, and fitting better around the excited state, thus lowering its energy. The net result is a lowering of the transition energy.

Polymer Spectra

The CD spectrum of HA polymer in aqueous solution at pH 2.5 is shown in Figure II.3. It consists of negative CD in the n-π^* transition region near 210 nm and an approach to the baseline in the π-π^* region near 190 nm, with negative CD at shorter wavelengths. Thin cell measurements at room temperature show that the second negative CD band in HA is centered at a wavelength shorter than 173 nm, and is at least 1.5 times the intensity of the 210 nm band. Surprisingly, there is little CD in the π-π^* region, considering that

Figure II.3 CD spectrum of HA. The CD spectrum of HA was recorded in aqueous solution buffered at pH 2.5 with 20 mM sodium phosphate (———). For comparison, a spectrum which is the sum of the component monomer spectra is also presented (— — —).

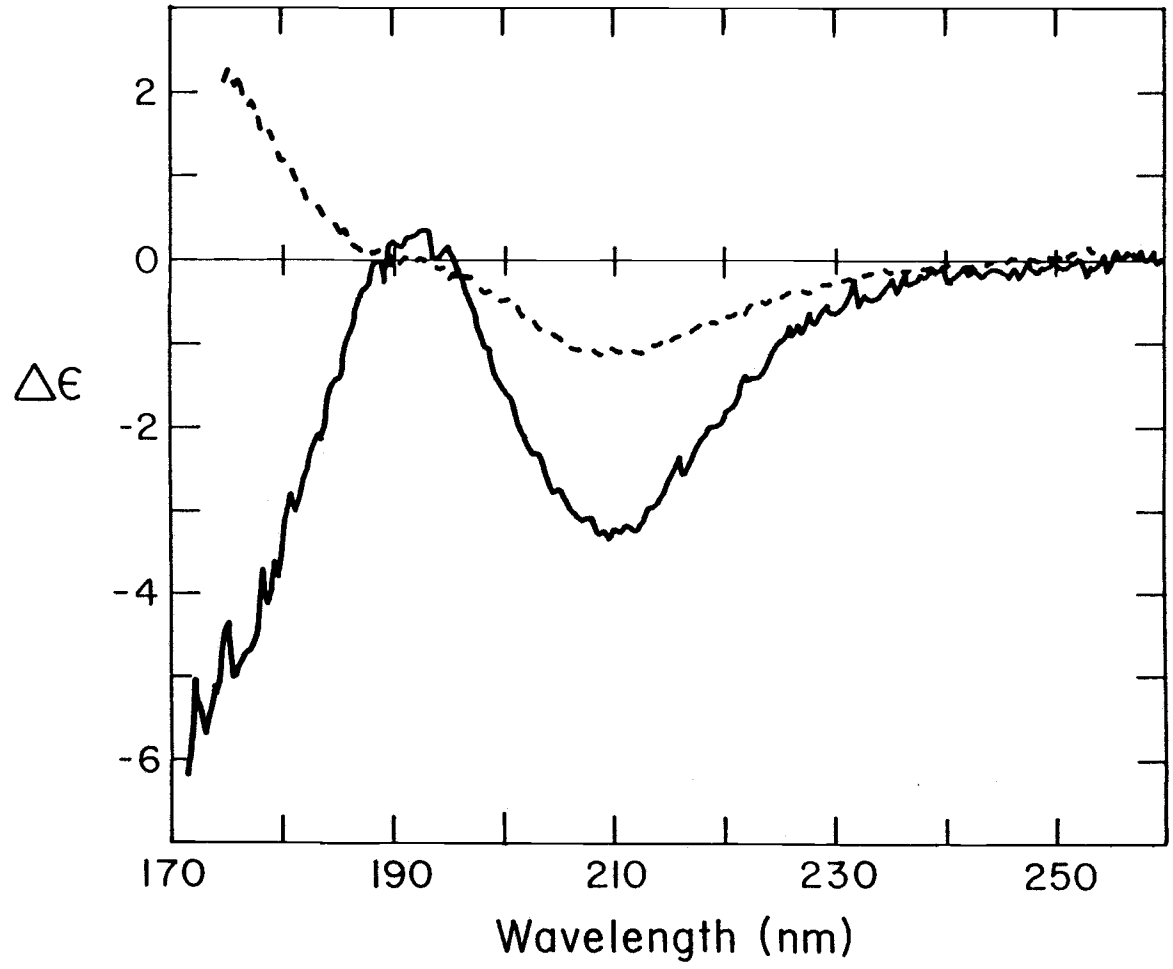

Figure II.3

both the acetamido and the carboxyl groups of HA should have transitions there. Buffington et al. (1977) proposed that cancellation of monomer transitions resulted in the weak signal of HA at 190 nm. In comparing monomer and polymer spectra, they were able to show that the changes in HA CD resulting from an increase in pH are paralleled by similar changes in the CD of GlcUA. Subtle changes as a function of pH also occur in the π-π^* transition region of absorption and flow linear dichroism (LD) spectra of HA.

Flow linear dichroism measurements for the normal absorption of the polymer at pH 6.8 and 2.5 reveal signal changes related either directly to protonation of the carboxyl group, or to changes in orientation of the carboxyl and acetamido groups relative to the direction of flow. At pH 6.8, the flow LD spectrum contains a broad, trapezoidal shaped and negative band from 175 to 190 nm which appears to contain two transitions (figure II.4). With acidification, the total signal intensity decreases, consistent with reduction of the persistence length of the molecule as electrostatic repulsion of the negative carboxyl groups is alleviated. The shape of the LD signal also changes, indicating the presence of distinct transitions at wavelengths of 187 nm and shorter than 180 nm. The intensity of the 180 nm band is reduced relative to that of the 187 nm band at low pH, whereas they appear to be almost equal at pH 6.8.

These results, when combined with the presence of absorption maxima for HA and GlcNAc at 188 nm and the absence of such a maximum in the corresponding spectrum of GlcUA, suggest that the π-π^* transition of GlcNAc in HA lies near 188 nm, with the π-π^* transition of GlcUA occurring at shorter wavelength. The overlap of transition

Figure II.4 Flow LD of HA. Flow LD spectrum of HA in aqueous solution was recorded at pH 2.5 (— — —) and 6.8 (———), 24 degrees C. HA concentration was 1 mM in disaccharide.

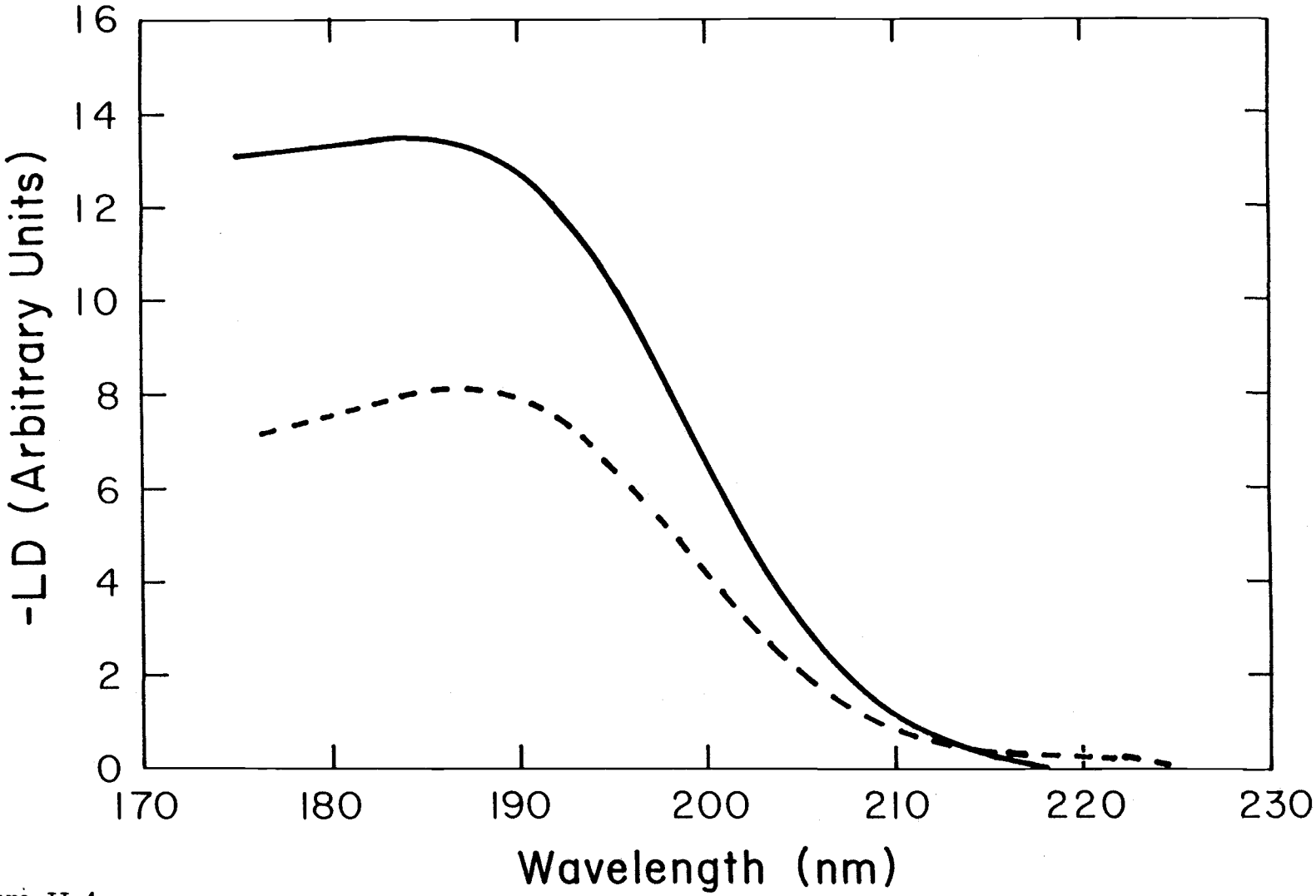
16
14
12
10
8
6
4
2
0
-LD (Arbitrary Units)
170
180
190
200
210
220
230
Wavelength (nm)

Figure II.4

bands within and between monomer spectra makes precise assignments in the HA spectrum impossible, especially since transition energies and intensities may change upon polymerization of the monomers. The work of Stone (1969, 1971) suggests that the transitions of the acetamido group dominate the optical activity of glycosaminoglycans generally, since these polymers all have a negative CD band in the $n-\pi^*$ transition region also found in amino sugar derivatives, and a second band in the $\pi-\pi^*$ region whose sign is dependent on the type of backbone linkage to the amino sugars. This dominance may not remain in nonaqueous solution, however.

For comparison with the HA polymer spectrum, a sum of monomer spectra has been included as the dashed line in Figure II.3. Clear differences exist between the two curves. The polymer has far more negative CD than the monomer sum in the $n-\pi^*$ transition region. At wavelengths shorter than 190 nm the two spectra even differ in sign.

The difference between the HA polymer and the combined monomer spectra has been noted previously. Various origins for this difference have been discussed, including conformational order within the polymer backbone (Chakrabarti and Balazs, 1973), changes in average chromophore orientations in monomers as affected by neighboring residues (Buffington et al., 1977), and the conversion of hemiacetal chromophores in the monosaccharides to acetal chromophores with glycosidic bond formation in the polymer (Cowman et al., 1983).

Cowman et al. (1981, 1983) have examined the origin of HA polymer CD by isolating oligomers of the molecule created by enzymatic cleavage of one of the two interior glycosidic linkages. Using an endo-β-hexosaminidase (bovine testicular hyaluronidase) to cleave at

the glycosidic bonds of the amino sugars, and an endo-β-glucuronidase (leach head hyaluronidase) to cleave at the glycosidic bonds of uronic acids, they have shown that the enhanced negative CD of HA polymer in the n-π* region is due to intact glycosidic bonds from C1 of GlcNAc to C4 of GlcUA. Thus as the number of these bonds per disaccharide increases with chain length in an oligomer series created by testicular hyaluronidase digestion, the negative CD at 210 nm also increases. The 210 nm CD in the corresponding oligomer series created by the leach head hyaluronidase is essentially constant (Cowman et al., 1981). Cowman et al. (1983) later extended this work into the π-π* transition region, where they showed the two interior linkages of HA to contribute CD of opposing signs. The hexosaminidic linkages contribute positive CD near 190 nm, and the glucuronidic linkages contribute negative CD. In the HA polymer, the two contributions approximately cancel. The resulting CD of HA as a function of chain length can be modeled as an end effect.

We will return to a discussion of chain length dependence of CD in Chapter IV, where we will present data describing an effect of chain length on HA CD distinct from an end effect. The end effect described by Cowman et al. (1981, 1983) is presented here as it likely accounts for the difference in CD between HA polymer and monomers in aqueous solution, whether under acidic condition or at neutrality.

The CD of the HA polymer in aqueous solution of pH 2.5 is shown in Figure II.5 as a function of temperature in the range of 0 to 60 degrees C. The decrease in CD magnitude of the 210 nm band assigned to the acetamido n-π* transition, and the decrease in positive CD in the π-π* transition region are approximately linear with temperature

Figure II.5 Temperature dependence of HA polymer CD. CD spectra of HA polymer dissolved in aqueous acidic solution, pH 2.5, recorded as a function of temperature. Temperatures in degrees C, -1 (———), 12 (— · · —), 20 (— — —), 32 (— · —), 47 (— —), 62 (· · · · · · · · ·).

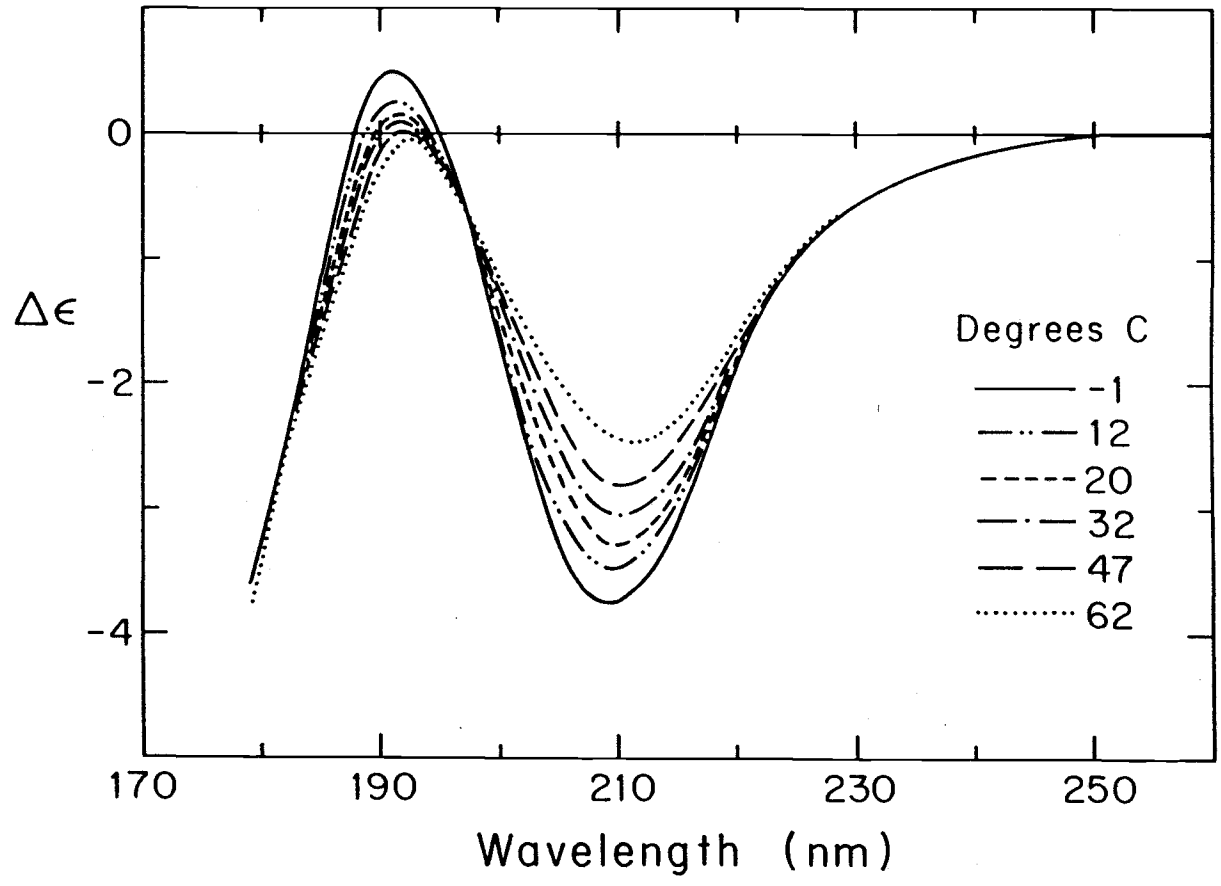
0
-2
-4
Δε
170
190
210
230
250
Wavelength (nm)
Degrees C
-1
12
20
32
47
62

Figure II.5

in this range. There is an apparent isodichroic point near $\Delta\varepsilon_{197}$ = -0.5. These CD changes with temperature are rather minor and perhaps uninteresting, but we must consider them as potential complications in data concerned with the melting of ordered structure in HA.

Unlike the subtle changes in CD observed as a function of temperature in aqueous solutions of HA, those changes occurring with introduction of an organic solvent to aqueous HA solution can be quite striking and are cooperative in nature. The change in the CD of HA polymer upon introduction of ethanol to an acidic aqueous solution is shown in Figure II.6. With addition of ethanol, appearance of strong negative CD at 188 nm is accompanied by loss of negative signal in the n-π^* transition region of 210 nm. The change is essentially complete at 25% ethanol, and the resulting spectrum is similar to that observed for NaHA films (Buffington et al., 1977) with a blue shift of the strong negative CD band from 194 nm in film to 188 nm in the acidic mixed solvent. Measurements of solutions with 2,2,2-trifluoroethanol substituted for ethanol have an isodichroic point near $\Delta\varepsilon_{176}$ = -5.0 in addition to the one at $\Delta\varepsilon_{204.5}$ = -2.7. Measurements to 169 nm using thin cell apparatus show a crossing of the baseline near 175 nm, with positive CD at shorter wavelengths. The presence of an isodichroic point demonstrates that the spectral series is composed of linear combinations of only two distinct species. This does not necessarily mean that any conformational transition of the polymer molecule as a whole, reflected in this CD change, is a two-state one. It does imply that those chromophores in HA contributing to the CD change with solvent experience one of only two environments.

Figure II.8 presents results of mathematically decomposing 16 CD

Figure II.6 Dependence of HA CD on ethanol concentration. CD spectra at 25 degrees C for HA dissolved in solutions with various proportions of ethanol to aqueous phosphoric acid solution, pH 2.5. The polysaccharide was lyophilized after removal of all cations by exchange resin AG50W (Bio Rad). It was then dissolved in water and dialyzed four times against a 500-fold excess of the desired solvent. Ethanol proportions as a percentage volume/volume of total solution were 0 (– – – –), 8.2 (— —), 10.2 (— · —), 14.6 (— – —), 18.8 (— – – —), and 45.9 (———). Similar results were obtained by gravimetric mixing of samples of polymer in acidic aqueous solution with trifluoroethanol. The isodichroic point at 205 nm identifies a two component system. Dotted curves are reconstructions of the sample spectra using only the two most significant components of an orthogonal basis set derived from these spectra and ten others (Figure II.8).

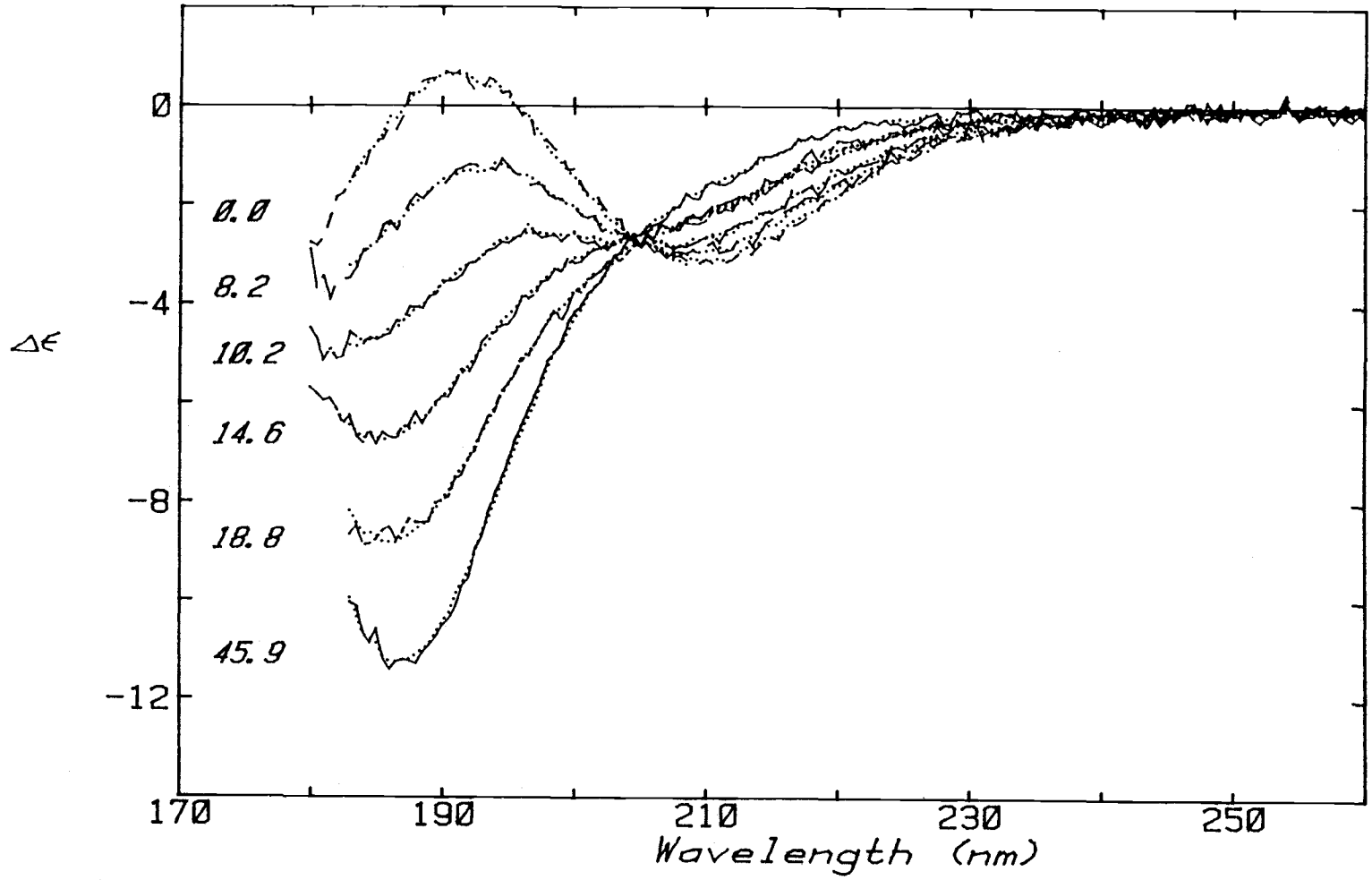
0
-4
-8
-12
Δε
0.0
8.2
10.2
14.6
18.8
45.9
170
190
210
230
250
Wavelength (nm)

Figure II.6

Figure II.7 CD of HA in solutions containing 2,2,2-trifluoroethanol. Substitution of trifluoroethanol for ethanol allowed further penetration of the vacuum ultraviolet region, revealing a second isodichroic point in spectra of varying solvent composition, near 176 nm.

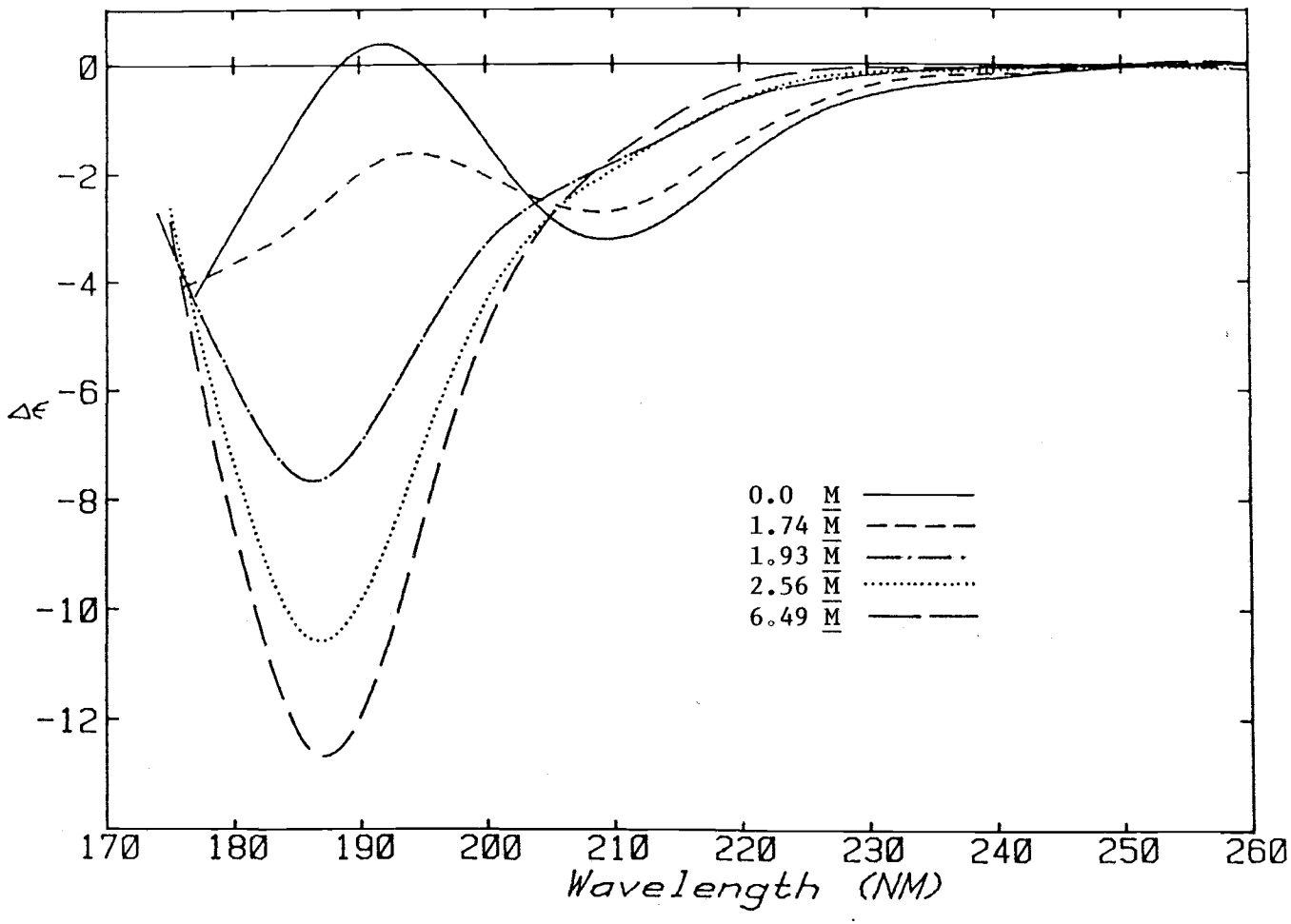
0
-2
-4
-6
-8
-10
-12
Δε
170
180
190
200
210
220
230
240
250
260
Wavelength (NM)
0.0 M
1.74 M
1.93 M
2.56 M
6.49 M

Figure II.7

Figure II.8 Orthogonalization of HA CD spectra recorded as a function of ethanol concentration of the solvent. 16 spectra of HA solutions with various alcohol contents, including those of Figure II.6, were analyzed for orthogonal components using singular value decomposition (Hennessey and Johnson, 1981; Forsythe et al., 1977). The four most significant components are presented, with their corresponding eigenvalues. All of the useful spectral information is contained in the two components with the largest eigenvalues.

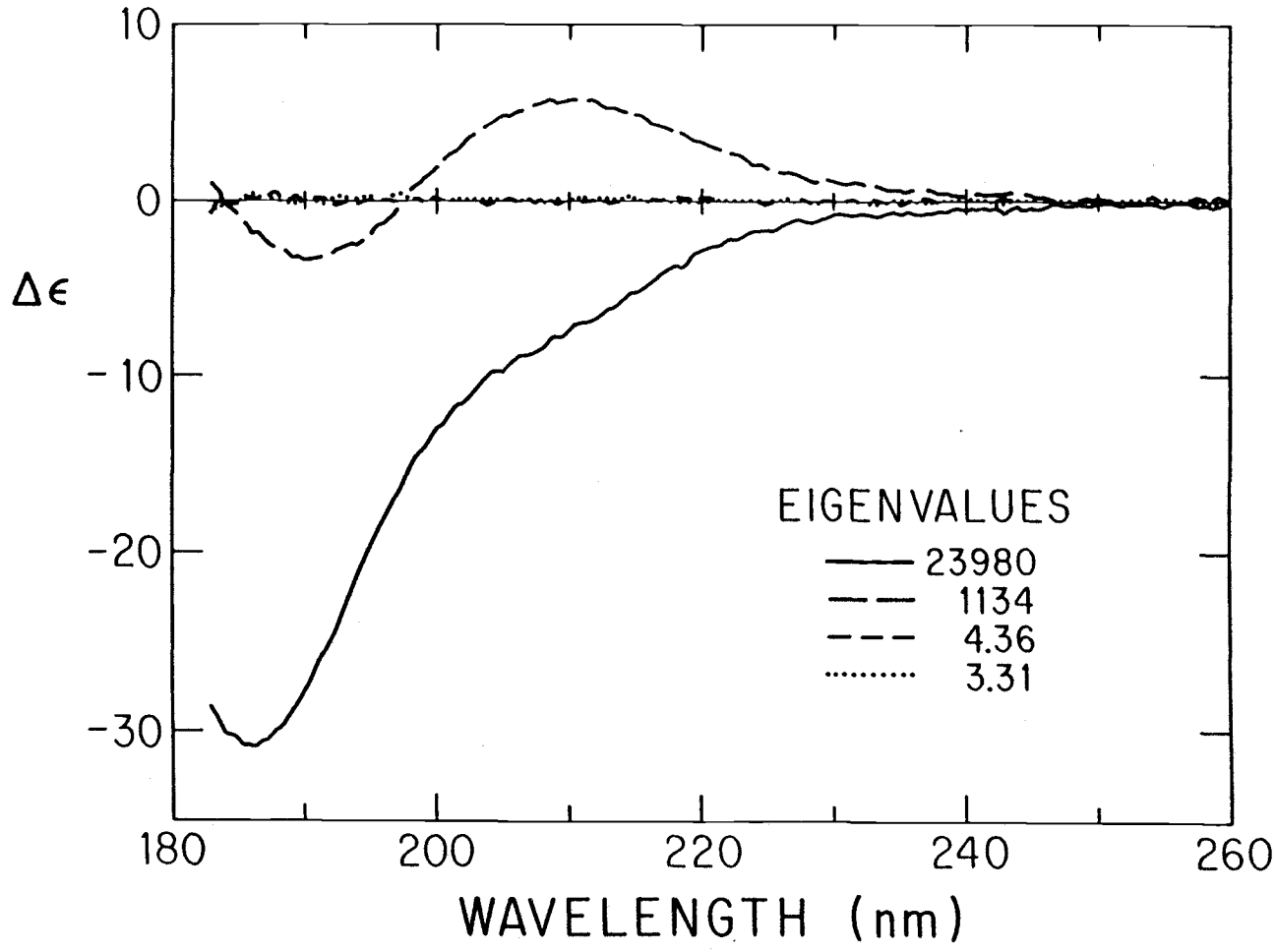
10
0
-10
-20
-30
Δε
180
200
220
240
260
WAVELENGTH (nm)
EIGENVALUES
23980
1134
4.36
3.31

Figure II.8

spectra of HA solutions with various contents of alcohol into orthogonal components following Hennessey and Johnson (1981). Only the four most significant components, as judged by the magnitudes of their corresponding eigenvalues, are presented. Reconstructions of the spectra in Figure II.6 using only the two dominant members of the orthogonal set, show that these two basis spectra contain all of the CD information within the noise level of the measurements, verifying the expectation from the presence of isodichroic points.

Park and Chakrabarti (1977a) were the first to observe a chiroptical transition of HA as a function of solvent composition. They described the cooperativity of the transition and its reversal with increase in pH or temperature. They also correlated the change in CD with a change in intrinsic viscosity of the polymer solution (Park and Chakrabarti, 1978b). Our results differ in several important respects from those of Park and Chakrabarti. Those authors utilized the wavelength region near 225 nm to describe this chiroptical transition, but shorter wavelengths are much more sensitive to the transition. We have extended the CD measurements of HA into the vacuum UV region to clearly show the negative CD band maximum at 187 nm which increases to a negative $\Delta\varepsilon_{188}$ more intense than -11 $\underline{M}^{-1}cm^{-1}$ at 46% v/v ethanol. Figure II.9 of $\Delta\varepsilon_{190}$ vs ethanol content of the solvent shows the cooperative nature of the CD as a function of solvent composition. The midpoint for the change is approximately 12.5% v/v ethanol at 24 degrees C and 2-3 m$\underline{M}$ concentration of disaccharide. The transition is essentially complete at 25% v/v ethanol. These results are in agreement with those of Park and Chakrabarti obtained by monitoring a chiroptical transition at 225

Figure II.9 HA $\Delta\varepsilon_{190}$ versus ethanol concentration. Most of the chiroptical transition is completed at 25% v/v ethanol. For deionized polymer sample, the ethanol content of the solvent could be increased to exceed 50% v/v without spectroscopic evidence of sample aggregation. The midpoint in terms of alcohol content for the transition at 25 degrees C is approximately 12.5%. Spectra were not sensitive to HA polymer concentration in the range of 0.5 to 10 mM in disaccharide. The vertical bars represent the noise level in CD measurements.

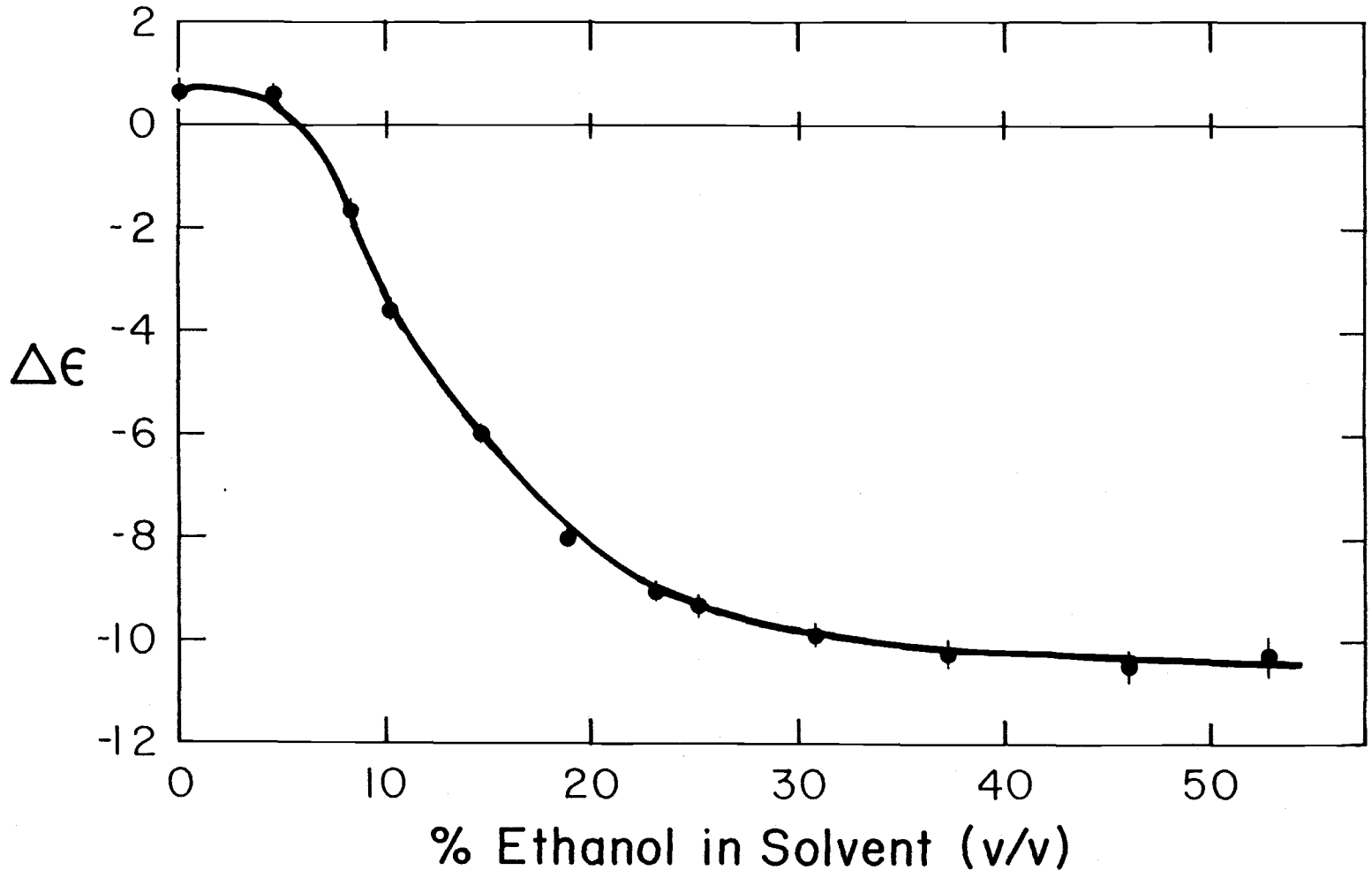

Figure II.9

nm. However, we do not observe positive CD in the 225 nm region as presented by Park and Chakrabarti (1978a,b), either for oligomers of HA in buffered mixed solvent (see Chapter IV) or for deionized polymer in acidic mixed solvent. It is important in this regard that our samples of HA polymer were deionized by cation exchange to remove metal counterions. We have observed positive CD signals in spectra of HA polymer, but only in the presence of monovalent counterion, which at low pH induces polymer gel formation at the concentrations used in our experiments. Such spectral results are *not* reliable due to linear dichroism (LD) contributions to the CD. We suggest that the positive CD at 225 nm observed by Park and Chakrabarti results either from sample orientation, becoming more pronounced with polymer aggregation in the presence of counterion, or from differential scattering of the circularly polarized light by aggregated material (Bustamante et al., 1983). It is noteworthy that Park and Chakrabarti (1978a) describe a concentration dependence of the positive CD at 225 nm, whereas we observe no such dependence in our spectra of HA polymer. If the positive CD at 225 nm is a result of aggregation of sample material, it would be anticipated to display sensitivity to sample concentration.

Park and Chakrabarti do not present data for HA solutions with ethanol concentrations higher than ~25% v/v. This may well represent a practical upper limit to the alcohol content achievable in their solutions without gelation of the sample. Using deionized HA polymer, we find that we can increase the ethanol content of our solutions in excess of 50% v/v without aggregation of the HA. In the process, we observe small CD changes when the alcohol content is increased beyond

25% v/v. In subsequent work we utilize an ethanol concentration of 45% v/v, as it appears to represent a true endpoint in the CD changes observed as a function of ethanol concentration, without the complicating factor of polymer aggregation.

We have made several attempts to measure the CD spectrum of unoriented polymer gel in 45% v/v ethanolic solvent, pH 3, containing 20 mM sodium ion. These attempts included heating the sample to melt the gel while introducing it into the CD cell, squashing the gel material between quartz windows in a pseudo random fashion, and adding an aliquot of concentrated buffer solution to deionized polymer in acidic mixed solvent within a cell. This last approach was the most effective for achieving a CD spectrum of a sample with minimal orientation as judged by LD. The CD of such samples at 188 nm was approximately -15 $M^{-1}cm^{-1}$ at 25 degrees C. This value is considerably smaller than the estimate of -21 at 189 nm by Park and Chakrabarti (1978a) for deionized HA in 20% ethanol solution at pH 3. Our estimate of the CD at 188 nm for HA gel rests on the assumption that the monovalent ions added to the sample material diffused uniformly throughout the cell. In Chapter IV evidence from oligomer studies will be presented which is consistent with this estimate. The CD determined by Park and Chakrabarti at 189 nm may not be reliable for a different reason. Those authors were utilizing 1 cm pathlength cells, in which absorption due to the solvent alone will exceed 1 O.D. at 190 nm. Their measurement must therefore be considered sensitive to any stray light artifact present in the spectrometer.

The changes in HA CD observed with increasing temperature in acidic mixed solvent (Figure II.10) are essentially a reversal of

Figure II.10 Melting of ordered HA structure in 33% v/v acetonitrile solution. CD spectra were recorded at various temperatures in an acidic aqueous solution containing 6.34 M acetonitrile. Temperatures in degrees C, 2.3 (———), 22.3 (— - - —), 30 (— — —), 35 (— · —), 38 (— —), 40 (— — —), 45 (·····), 64 (———). The CD spectrum at 15 degrees C is only slightly different from 2.3 degrees C.

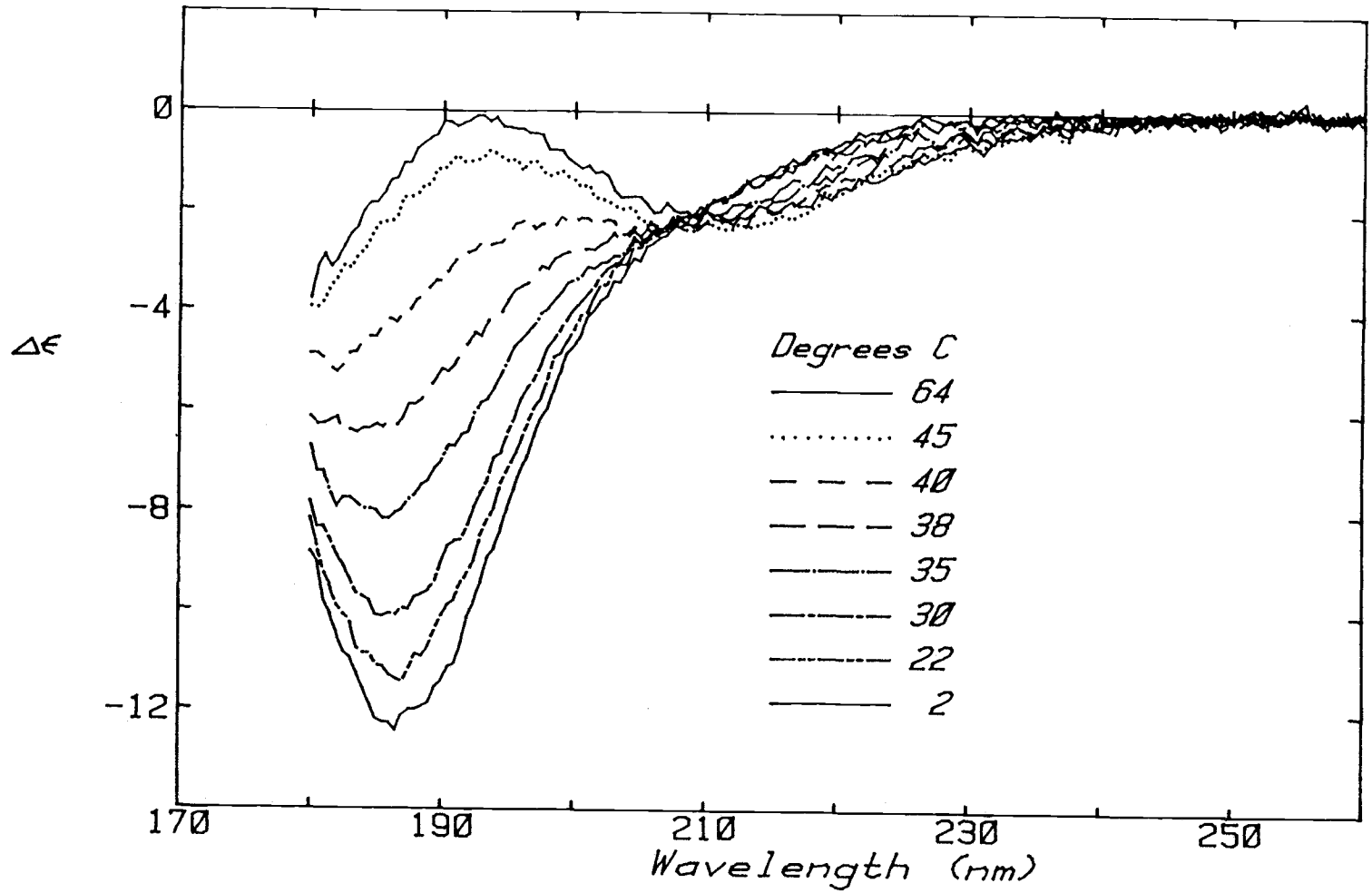
0
-4
-8
-12
Δε
170
190
210
230
250
Wavelength (nm)
Degrees C
64
45
40
38
35
30
22
2

Figure II.10

those resulting from addition of ethanol to an acidic aqueous solution of HA as depicted in Figure II.4. With increasing temperature the negative CD in the π-π^* transition region of 188 nm is lost, with a concomitant increase of negative CD in the n-π^* transition region at 210 nm.

A series of spectra recorded during melting of the ordered structure of HA in acidic aqueous-organic solvent is presented in Figure II.10. The solvent was 6.34 M in acetonitrile. Previous HA structure melts performed in solutions containing alcohol were not freely reversible after exposure of the sample to temperatures near 100 degrees C for several hours. Since this may be due to esterification of HA carboxyl groups at high temperature and low pH, acetonitrile was substituted for alcohol in these measurements. Unlike the situation with varying solvent composition, there is no clear isodichroic point for this series of spectra, and reconstructions of the original spectra within the noise level of the measurements require at least three basis components. This is not unexpected, since the CD of HA in acidic aqueous solution, which may represent the CD of unordered HA in both solvents, is observed to change with temperature in a noncooperative manner (see Figure II.5). Above 60 degrees C, the CD spectrum of HA in acidic mixed solvent is nearly identical to its spectrum in aqueous solution at that temperature (compare figure II.10 with figure II.5). This identity supports our supposition that HA is essentially unordered in aqueous solution, and that introduction of organic solvent has little effect on the CD of unordered HA. The small differences between these spectra in band amplitudes and frequencies are similar in nature to

those spectral changes observed in the monomer (i.e. GlcNAc and GlcUA) CD upon introduction of organic solvent.

The CD at 190 nm for HA in the water-acetonitrile solvent system is plotted as a function of temperature in Figure II.11. The wavelength was chosen for its proximity to the negative CD maximum at low temperature (187 nm), and the low intensity of CD observed in aqueous HA solutions at 190 nm which is presumably due to the π-π^* transition. The dashed line of Figure II.11 is the minor change in HA CD observed as a function of temperature at 190 nm in acidic aqueous solution.

Using our $\Delta\varepsilon_{188}$ for ordered HA of -12 $\underline{M}^{-1}cm^{-1}$, and the fact that the CD of unordered HA in acidic mixed solvent is the same as that of unordered HA in acidic aqueous solution, data such as those in Figure II.11 can be analyzed in terms of a simple two-state transition theory. Briefly, the CD at 188 nm can be used to estimate the fractions of ordered and unordered HA polymer at various temperatures. From these values, an equilibrium constant K for the transition between ordered and unordered species can be calculated using one of various models involving an intra- or inter- strand reaction described in Chapter III. A plot of the logarithm of K versus 1/T, where T is the absolute temperature, will provide an estimate of the transition enthalpy from the slope of the graph following the van't Hoff relation. Such a plot is shown in Figure II.12, in which K corresponds to a bimolecular association. Only data greater than 10% of the total transition range from both endpoints are plotted. The van't Hoff plot is nonlinear, due to temperature dependence of the transition enthalpy, failure of the simple model to describe the

Figure II.11 CD melting curve at 190 nm for HA in water-acetonitrile solution. $\Delta\varepsilon_{190}$ is plotted as a function of temperature for HA in an acidic aqueous solution containing 33% v/v acetonitrile. The dashed line is 190 nm CD for HA in acidic aqueous solution. Vertical bars represent noise level in the measurements.

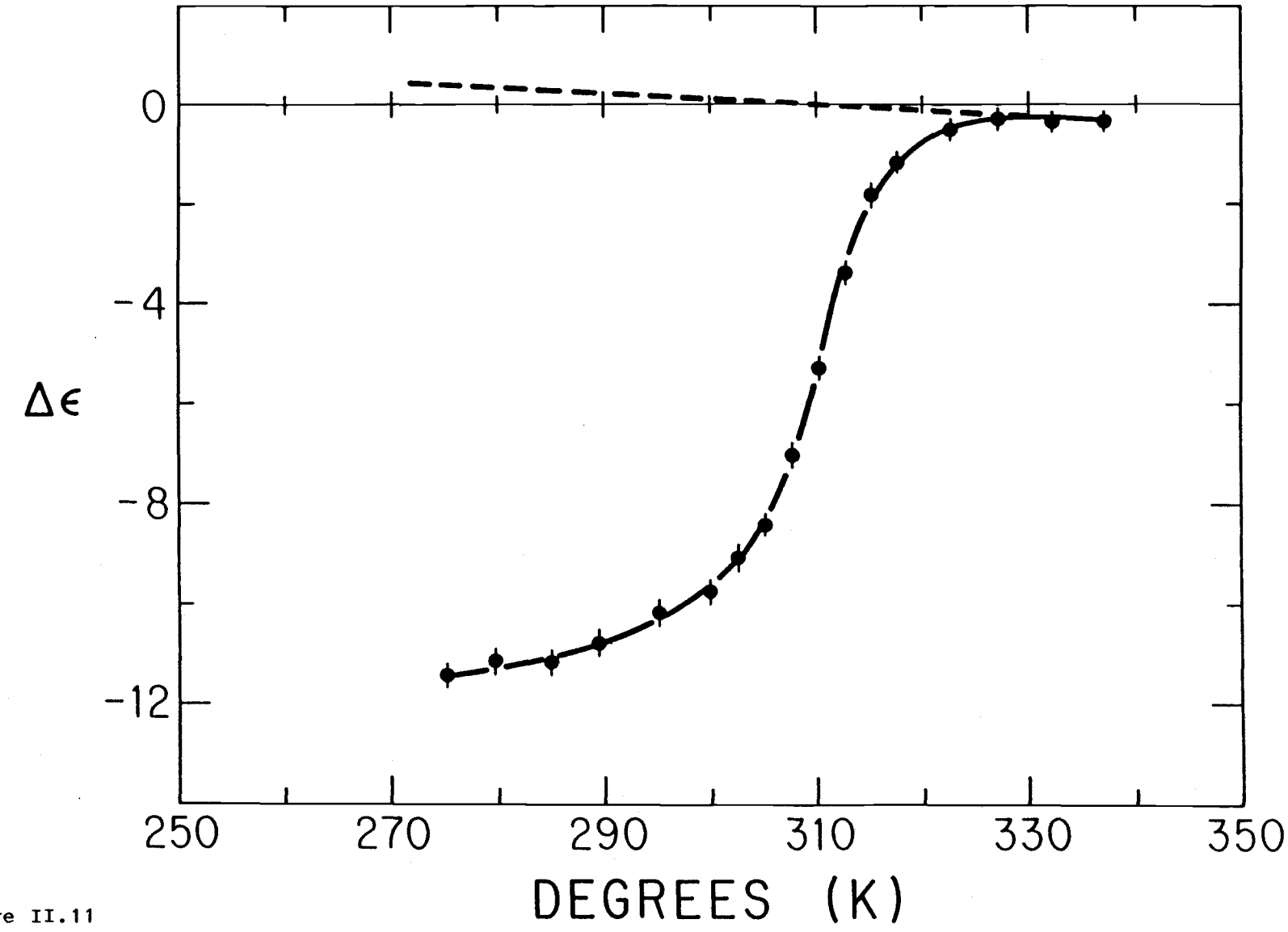

Figure II.11

transition, or both.

There is no evidence of concentration dependence in the CD of HA polymer measured here. However, the CD data described in Chapter IV for oligomers of HA do show a concentration dependence. Thus analyses of the data using models of two or three strand association can be attempted, assuming that the molecular weights of the molecules are known so that HA strand concentrations can be calculated. We have analyzed the data of Figure II.11 assuming two-state associations of two or more strands in calculating equilibrium constants, and using extremes of 1 or 400 disaccharides as the chain lengths of the HA strands.

We will perform linear regression analysis on the data in plots such as Figure II.12, and use the correlation coefficient to judge the linearity of the relationship described by the data. When we compare models involving different orders of strand association in this way, certain trends become apparent. The higher order association models provide more linear estimates of ℓnK versus 1/T. This is more pronounced in comparing uni- with bimolecular, and bi- with trimolecular models. The difference in linearity between tri- and tetramolecular associations is so small as not to be discernible except in the correlation coefficient. Despite the fact that plots of ℓnK vs 1/T become more linear as the order of association in the model is increased, the error still appears non random. The values of ℓnK consistently fall below the line at high T, then cross above it for intermediate T, and cross the line once again at low T. They describe a gentle 'S' along the linear estimate, as in figure II.12.

The estimate of HA chain length, reflected in total strand

Figure II.12 van't Hoff plot of HA in water-acetonitrile solution. Data of Figure II.6 were analyzed in terms of a two state transition model. $\Delta\varepsilon_{188}$ for the low temperature limit of the transition was assumed to be -11.8 $\underline{M}^{-1}cm^{-1}$ (independent of temperature), and the high temperature limit was assumed to be the same as that in acidic aqueous solution (dashed line in Figure II.11). The equilibrium constant was calculated at various temperatures for the reaction 2 MONOMER → DIMER, where the dimer is the low temperature species. Assuming a unimolecular transition equilibrium constant did not improve the linearity of the plot. Only those data at a distance from either endpoint which was at least 10% of the total transition range are plotted. Data rectangle height and width correspond to CD noise level and uncertainty in temperature, respectively.

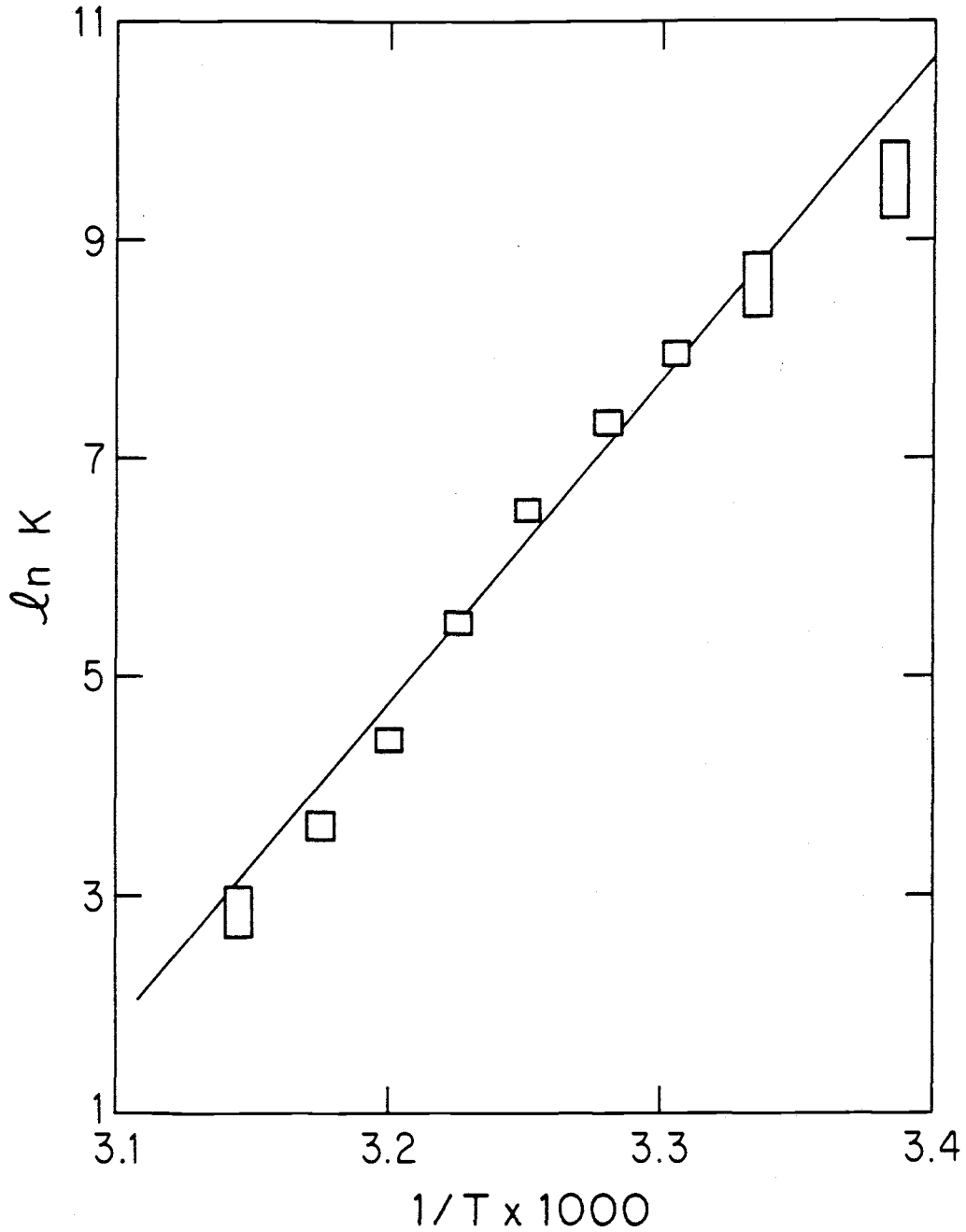

Figure II.12

concentration, has little influence on the results of model fitting. Whether strands are assumed to be one or 400 disaccharides long, changes in the slope of a ℓnK vs $1/T$ plot for a bimolecular association differ by less than one part in 350, from 351.6 to 352.2. A larger influence is exerted by the estimate of $\Delta\varepsilon_{188}$. The linearity of van't Hoff plots for all models examined increases as $\Delta\varepsilon_{188}$ is reduced from -13 $\underline{M}^{-1}cm^{-1}$ to -12 $\underline{M}^{-1}cm^{-1}$. This will be confirmed in Chapter IV where results of analyzing HA oligomer transitions with cooperative as well as two-state models are presented.

The sigmoidal nature of CD changes for HA solutions as a function of solvent composition or temperature suggests that a cooperative transition model is more appropriate for this system. In synthetic polypeptide (Hayashi et al., 1969; Zimm et al., 1959; Gaskin and Yang, 1971; Nakamoto et al., 1974; Shaw and Schurr, 1975) and polynucleotide (Applequist and Damle, 1965; Martin et al., 1971; Nelson et al., 1981) transition studies, the use of oligomeric species has proven useful in understanding the transition process and describing the thermodynamic parameters involved. In Chapter IV we will show that CD measurements of various HA oligomers in the acidic mixed solvent system reveal not only the type of chain length dependence expected in a cooperative polymer transition, but also the dependence on sample concentration expected for transitions involving strand association. In fact, data describing oligomer transitions as functions of temperature, strand length, and strand concentration can be analyzed to determine the most appropriate model and its associated thermodynamic parameters. We will demonstrate such an analytical procedure for the present transition after a discussion of the theoretical background for

cooperative systems in Chapter III.

CHAPTER III: TRANSITION THEORY

INTRAMOLECULAR TRANSITIONS

Consider as an example a system composed of Z polymer molecules, each containing n residues of a single type. Let the molecules be capable of an intramolecular transition. Assume that each residue of a polymer molecule can exist in one of only two states, which will be designated an ordered or helical (h) state, and an unordered or coiling (c) state. Assuming that the molecules do not interact with one another, the partition function for the system can be written

$$Q_{sys} = \frac{Q_{mol}{}^{Z}}{Z!} \qquad \text{(III.1)}$$

where Q_{mol} is the partition function for a single molecule. What is the form of Q_{mol}? Q_{mol} is a sum over Boltzmann terms for the ℓ states accessible to a single molecule, each of energy E_ℓ,

$$Q_{mol} = \sum_{\ell} e^{-E_\ell/kT} \qquad \text{(III.2)}$$

where T is the absolute temperature and k is the Boltzmann constant.

Now the Boltzmann term corresponding to state ℓ of the molecule is proportional to the probability $p(\ell)$ that the molecule exists in state ℓ with the corresponding energy E_ℓ (see appendix I). The relative probability of two molecular states of different energy can be expressed as the ratio of their Boltzmann terms. The molecule must

exist in some state, and so the sum of Boltzmann terms, or partition function, serves as a normalizing factor to provide absolute probabilities $p(\ell)$.

$$p(\ell) = \frac{e^{-E_\ell/kT}}{\sum_m e^{-E_m/kT}} = \frac{e^{-E_\ell/kT}}{Q_{mol}} \qquad \text{(III.3)}$$

If more than one molecular state can have energy E_ℓ (degeneracy), the expression for Q_{mol} can be changed accordingly to a summation over energy levels, as opposed to a summation over states.

$$p(\ell) = \frac{d_\ell e^{-E_\ell/kT}}{\sum_m d_m e^{-E_m/kT}} \qquad \text{(III.4)}$$

where d_ℓ is the number of states with energy E_ℓ, $p(\ell)$ is now the probability that the molecule has energy E_ℓ.

To discuss the relative probabilities of the helical and coiling states we establish a reference state for the molecule with all n residues in the coiling form, and define its energy as E_o. Then dividing Q_{mol} by the reference Boltzmann term,

$$\frac{Q_{mol}}{e^{-E_o/kT}} = 1 + \sum_{m\neq 0} \frac{d_m e^{-E_m/kT}}{e^{-E_o/kT}} = 1 + \sum_{m\neq 0} d_m e^{-(E_m-E_o)/kT}$$

$$= 1 + \sum_{m\neq 0} d_m e^{-\Delta E_m/kT} = 1 + \sum_{m\neq 0} d_m q_m \qquad \text{(III.5)}$$

where $q_m = e^{-\Delta E_m/kT}$.

Note that $kT\ell nq_m = -\Delta E_m$. Thus q_m is a microscopic equilibrium constant in terms of molecules, for the formation of a polymer molecule with energy E_m from one with energy E_o (Cantor and Schimmel, 1980, p. 1054). It is the ratio of the number of molecules with energy E_m (proportional to the Boltzmann term), to the number of molecules with energy E_o.

The Noncooperative Case

Let q_m be such an equilibrium constant for the formation of an h residue from a c residue in the polymer molecule, and assume that q_m is independent of the residue's location in the polymer chain. Then if the residues are independent, the partition function for the molecule is simply a binomial sum of products of q_m and 1, with a q_m occurring for each h residue and a 1 for each c residue (Cantor and Schimmel, 1980, p. 1055).

$$\frac{Q_{mol}}{e^{-E_o/kT}} = (q_m+1)^n = \sum_{i=o}^{n} \frac{n!}{(n-i)!i!} q_m^{n-i} \qquad \text{(III.6)}$$

The quantity n is the degree of polymerization.

Zipper Model

Cooperative polymer transition models use two separate energies in the transition process. One energy may be regarded as the energy for initiation of a conformational change in a polymer of

conformationally uniform residues, with the initiation taking place at a single one of those residues. The other energy may be regarded as the energy for propagation of the conformational change through the rest of the polymer. It is the energy for conformational change of a residue at the interface between regions of differing conformation in the polymer. We shall define this latter energy to be $-kT\ell ns$, where s is the equilibrium constant for the transition of a residue in a polymer molecule from one conformation to another, and the residue in question is at the interface between regions of the two conformations. The former energy for initiation of conformational change within a uniform region of residues shall be defined as $-kT\ell n\alpha s$. The α in the equilibrium constant αs is necessary because this change is occurring not at an interface between two regions of differing conformation, but within a conformationally uniform region. Therefore it may be expected to be different from s. This is distinct from the noncooperative case, in which a single equilibrium constant (q_m) serves for both conformational changes.

The s is present in αs to remind us that many aspects of the transition of a residue at initiation may also occur with propagation, as in the formation of α-helical structure in polypeptides. With initiation of α helix formation, a hydrogen bond is formed between amino acid residues four units apart in the polymer. Growth of the α helix by one residue at an end also results in formation of such a bond. Yet the two processes differ in the entropies involved. Initiation of the helix is accompanied by conformational ordering not only of the residues which are hydrogen bonded to one another, but also of the three residues between them. Growth of the helix requires

conformational ordering of just one residue at the end of the helix. The participation of more than one microscopic process in the conformational ordering of a polymer creates cooperativity in the transition. The initiation equilibrium constant αs is typically smaller than s, corresponding to a difference between the larger energy required for initiation of a helical region ($-kT\ell n\alpha s$), and the smaller energy for growth of an existing helical region by one residue ($-kT\ell ns$).

A simple model for cooperativity in the polymer transition which is particularly suitable for short chains is called the zipper model (Cantor and Schimmel, 1980, p. 1056). It allows for only one helical region in a molecule, ranging in size from one to n residues. A string of h helical residues can be located in a polymer of n total residues in n-h+1 ways. Then from the last equality in equation (III.5)

$$\frac{Q_{mol}}{e^{-E_o/kT}} \text{ becomes } 1 + \sum_{h=1}^{n} (n-h+1)\alpha s^h \qquad \text{(III.7)}$$

where $q_h = \alpha s^h$, $d_h = n-h+1$

Now what is the fraction of helical residues in the polymer molecule? Remember that the probability of a state (with h helical residues) is its Boltzmann term divided by the partition function. From equation (III.7)

$$p(h) = \frac{e^{-E_h/kT}}{Q_{mol}} = \frac{e^{-E_h/kT} \Big/ e^{-E_o/kT}}{Q_{mol} \Big/ e^{-E_o/kT}} = \frac{q_h}{Q_{mol} \Big/ e^{-E_o/kT}}$$

$$p(h) = \frac{\alpha s^h}{1 + \sum_{h=1}^{n} (n-h+1)\ \alpha s^h} \qquad \text{(III.8)}$$

Since h is defined to be the number of helical residues, and there are n-h+1 different states each with h helical residues, the fraction of helical residues θ is

$$\theta = \frac{\sum_{h=1}^{n} h(n-h+1)\ \alpha s^h}{\frac{Q_{mol}}{e^{-E_o/kT}} \times n} \qquad \text{(III.9)}$$

If we define $q_{mol} = \frac{Q_{mol}}{e^{-E_o/kT}}$ which is $1 + \sum_{h=1}^{n} (n-h+1)\alpha s^h$

then $\theta = \frac{s(\partial q_{mol}/\partial s)}{q_{mol} \times n} = \frac{1}{n} \frac{\partial(\ln q_{mol})}{\partial(\ln s)}$ (III.10)

INTERMOLECULAR TRANSITIONS

The data in this study were observed to be concentration dependent, with two significant components being found in the series of CD spectra measured as a function of concentration. It is presumed that the change in CD with concentration of oligomer reflects strand association, even if indirectly through some coincident secondary structural change. We want to analyze the data employing models allowing for molecular association of two or more strands. The observed residue CD (CD_{obs}) is therefore treated as a weighted sum of contributions from unassociated (free) residues (CD_{fr}), and associated (complexed) residues (CD_{com}) found in strand complexes. We will predict the fraction of free and complexed residues in the system using trial transition energies in the various models, and compare the calculated results with experiment.

$$CD_{calc} = \theta(CD_{com}) + (1-\theta)(CD_{fr}) \qquad \text{(III.11)}$$

Applequist and Damle (1965) demonstrated that transitions in associating polymers can be analyzed by considering the partition function for the entire system, as opposed to that for a single molecule. In a system of Z polymer molecules, each n residues long, in which two single molecules may associate to form a duplex, the system partition function will be

$$Q_{sys} = \sum_{i=0}^{Z/2} \frac{Q_{dup}{}^{i}}{i!} \frac{Q_{sin}{}^{(Z-2i)}}{(Z-2i)!} \qquad \text{(III.12)}$$

where Q_{dup} and Q_{sin} are the duplex and single molecule partition functions, and i is the number of molecular duplexes in the system. Because Z is quite large, Q_{sys} will be dominated by its largest term, which is also the term with the largest natural logarithm (Guggenheim, 1955, p. 56). Furthermore, from equation (III.12) the logarithm of this term will have the form

$$\ln Q_{sys}(i) \cong i\ln Q_{dup} - \ln i! + (Z-2i)\ln Q_{sin} - \ln(Z-2i)! \qquad \text{(III.13)}$$

To find the maximum value of $\ln Q_{sys}(i)$, we take the partial derivative with respect to i and set the expression equal to zero. Using the Stirling approximation $\ln N! = N\ln N - N$,

$$\frac{\partial \ln Q_{sys}(i)}{\partial i} = \ln \frac{Q_{dup}}{Q_{sin}^2} \frac{(Z-2i)^2}{i} = 0, \text{ and } \frac{Q_{dup}}{Q_{sin}^2} \frac{(Z-2i)^2}{i} = 1,$$

$$\text{or } \frac{Q_{dup}}{Q_{sin}^2} = \frac{i}{(Z-2i)^2} \qquad \text{(III.14)}$$

so $\frac{Q_{dup}}{Q_{sin}^2}$ is just the equilibrium constant for the association of two monomers to form a duplex written in terms of number of molecules.

Two-state Duplex

If Q_{dup} and Q_{sin} each have but one term (all n residue pairs in the duplex are in the helical conformation and all n residues in the monomer are in the coiling state), then the association involves a two-state transition in the sense that an entire molecule has only two available states. This all or none transition implies infinite

cooperativity between residues of the polymer chain as they switch from one conformation to the other. Due to this infinite cooperativity, the fraction of residues in the system that are associated is the same as the fraction of total oligomer molecules in the multi-strand complexes. θ in equation (III.11) above will depend on the concentration of the oligomer and the energy of the transition, through the equilibrium constant. In modeling a system with temperature and oligomer chain length as variables, we have divided the transition energy into length dependent and length independent parts. Each of these parts is allowed a temperature dependent (enthalpic) and independent (entropic) component.

The length dependent enthalpy and entropy are considered to be simple functions of the degree of polymerization, n, and are of the form $n\Delta H_{res}$ and $n\Delta S_{res}$. The symbol ΔH_{res} represents the contribution from a single residue to the enthalpy change in an oligomer transition. The corresponding entropy contribution is ΔS_{res}. The chain length independent enthalpy and entropy, ΔH_{ind} and ΔS_{ind}, are regarded as end effects so that the total transition energy becomes

$$n(\Delta H_{res} - T\Delta S_{res}) + \Delta H_{ind} - T\Delta S_{ind}. \qquad \text{(III.15)}$$

Our method of data analysis requires calculation of the fraction of residues in the system which are in the helical conformation. Since this model of the transition is a two-state one in which strands associate completely to achieve the helical conformation, this fraction is equivalent to the fraction of strands associated in the system. For a two strand association,

$$\theta = \frac{2[\text{complex}]}{[\text{total}]}, \quad 1 - \theta = \frac{[\text{free}]}{[\text{total}]} \qquad \text{(III.16)}$$

in terms of molar concentrations of free strands, complexes, and total strands. The equilibrium constant for the association is

$$K_{eq} = \frac{[\text{complex}]}{[\text{free}]^2} = \frac{\theta}{2[\text{total}](1-\theta)^2} \qquad \text{(III.17)}$$

and from the van't Hoff relation

$$-RT\ln K_{eq} = n(\Delta H_{res} - T\Delta S_{res}) + \Delta H_{ind} - T\Delta S_{ind} \qquad \text{(III.18)}$$

where n is the oligomer length, ΔH_{res} and ΔS_{res} are the enthalpy and entropy contributions per residue, and ΔH_{ind} and ΔS_{ind} are the corresponding chain length independent values. The equilibrium constant in terms of number of molecules ($q = Q_{dup}/Q_{sin}^2$) is related to K_{eq} which is in terms of molar concentrations through the system volume V and Avogadro's number N_o.

$$K_{eq} = qN_oV. \qquad \text{(III.19)}$$

Given the energy values and temperature, K_{eq} can be calculated from equation (III.18). Then θ can be determined from equation (III.17) above, and CD_{calc} can be calculated using equation (III.11) if CD_{fr} and CD_{com} are known.

Zipper Duplex

If the polymer association is not infinitely cooperative between residues, then Q_{dup} will have more than one term. Applequist and Damle (1963, 1965) simplify the calculation of Q_{dup} for short polymer chains by adapting the zipper model to an associating system in which strands associate in register and are parallel, and only one contiguous segment of residue pairs exists in a complex. We shall assume that residues attain a helical structure upon pairing in the complex. For one term in Q_{sin} (single molecules cannot contain helical residues) and using its energy as the reference energy, q can be written as a sum of microscopic equilibrium constants for the formation of the various duplexes from monomer molecules.

Cooperative duplex with parallel strands in register

In analogy to the intramolecular case, we can use Q_{sin}^2 to define the reference state and view Q_{dup} as Q_{mol}. Then from equation (III.7)

$$q = \frac{Q_{dup}}{Q_{sin}^2} = \sum_{h=1}^{n} (n-h+1) \frac{\alpha s^h}{V} \qquad \text{(III.20)}$$

where V is the system volume, and the first term is missing since Q_{dup} does not contain the reference state. Here αs is the equilibrium constant for formation of a duplex with a single helical residue pair from separated molecules. It is in terms of molecular concentrations. Symbol s is the equilibrium constant for addition of a helical residue pair to the end of a helical segment in a duplex. Expression n-h+1 is

the number of ways to arrange h contiguous helical residue pairs among n total residue pairs.

What is the fraction of helical residues in the system? It will be the average number of helical residues in each duplexed strand of the system ($\bar{h}$), times the number of strands in duplexes (2i), divided by the total number of residues in the system (Zn).

We shall continue the analogy to the intramolecular case presented in equations (III.7) through (III.10) to derive an expression for the average number of helical residue pairs in system duplexes. We will use our equilibrium constant for formation of duplexes from separated single strands, q of equation (III.20), and substitute its element for formation of a duplex with h helical residue <u>pairs</u>, q_h, for the corresponding intramolecular transition element used in the derivation of equation (III.8). The resulting probability of a duplex state with h helical residue pairs is

$$p(h) = \frac{\frac{\alpha}{V} s^h}{\sum_{h=1}^{n} (n-h+1) \frac{\alpha}{V} s^h} \qquad \text{(III.21)}$$

Since h is the number of helical residue pairs, and there are n-h+1 different duplex states each with h helical residue pairs, the average number of helical residue pairs in a duplex is

$$\bar{h} = \frac{\sum_{h=1}^{n} h(n-h+1)\frac{\alpha}{V}s^h}{\sum_{h=1}^{n} (n-h+1)\frac{\alpha}{V}s^h} = \frac{s}{q}\frac{\partial q}{\partial s} = \frac{\partial \ln q}{\partial \ln s} \qquad \text{(III.22)}$$

This is analogous to equation (III.9) defining the fraction of helical residues in an intramolecular transition. The n is missing from the denominator of equation (III.22) because we are considering the average number of residue pairs rather than the fraction of total residues which are paired. The fraction of helical residues will be

$$\theta = \frac{2i\bar{h}}{Zn} = \frac{2i}{Zn}\frac{s}{q}\frac{\partial q}{\partial s} \qquad \text{(III.23)}$$

The number of duplexes, i, can be replaced with q and Z by solving equation (III.14).

$$Z^2q - 4Zqi + 4qi^2 = i,\ 4qi^2 - (4Zq+1)i + Z^2q = 0$$

$$i = \frac{4Zq + 1 - \sqrt{8Zq + 1}}{8q} \qquad \text{(III.24)}$$

Substituting for i in equation (III.23), we have

$$\theta = \frac{4Zq + 1 - \sqrt{8Zq + 1}}{4Zqn}\frac{s}{q}\frac{\partial q}{\partial s} \qquad \text{(III.25)}$$

In practice, we want θ in molar rather than molecular terms. The number of strands in the system is related to their molar concentration, C_{tot}, by $Z/V=N_oC_{tot}$. Then

$$\theta = \frac{\left\{4C_{tot}N_o\alpha J(s) + 1 - \sqrt{8C_{tot}N_o\alpha J(s) + 1}\right\} s\partial J(s)/\partial s}{4C_{tot}N_o\alpha J(s)^2 n} \qquad \text{(III.26)}$$

where $J(s) = \sum_{h=1}^{n} (n - h + 1)s^h$. Furthermore

$$N_o\alpha = e^{-(\Delta H_{ind} - T\Delta S_{ind})/RT}, \quad s = e^{-(\Delta H_{res} - T\Delta S_{res})/RT} \qquad \text{(III.27)}$$

where the enthalpy and entropy will have trial values in molar units.

Cooperative duplex with antiparallel strands in register

The molecular equilibrium constant, q, for the duplex used in the preceding section was written presuming that the strands are parallel to one another (Applequist and Damle, 1965). If instead the two strands are antiparallel, duplexes with the same number of free residues at both ends will have a C2 symmetry axis perpendicular to the double helix axis and intersecting it midway along the length of the duplex. Other elements of the ensemble depicting all duplex states will lack this symmetry, but each such element will be related to another element of the ensemble through a 180 degree rotation about this axis. Still, each of these duplexes and its related element are not distinguishable in a quantum mechanical sense, since the strands are indistinguishable. Thus the molecular equilibrium constant, q, should be corrected to remove the redundancy (Kittel, 1969, p. 306). The factor Z! in the denominator of the partition function for a system of Z noninteracting molecules in equation (III.1) arises from this source. With an odd number of free residue pairs (n-h odd), none of the duplexes has C_2 symmetry, and all n-h+1 duplexes are redundant. With an even number of such pairs, one of the n-h+1 duplexes is not

redundant, and has C_2 symmetry. Removing the redundancy, the molecular equilibrium constant for the duplex becomes

$$q = \sum_{h=1}^{n} \left\{ \begin{array}{l} (n-h+1)\ \frac{\alpha}{2V}\ s^h,\ n-h\ \text{odd} \\ \\ (n-h+2)\ \frac{\alpha}{2V}\ s^h,\ n-h\ \text{even} \end{array} \right\} \qquad \text{(III.28)}$$

Relaxing the in-register condition

If the requirement that associated strands remain with their ends in register is relaxed, with the strands allowed to slide relative to one another in the complex, the number of states with energy $kT\ell n\alpha s^h$ (relative to the reference state) will increase. Now h contiguous residues arranged each of n-h+1 ways on one strand can associate with h residues arranged any of n-h+1 ways on the other strand. Applequist and Damle (1965) wrote the molecular equilibrium constant for this case with parallel strands as

$$q = \sum_{h=1}^{n} \frac{(n-h+1)^2}{2} \frac{\alpha}{V}\ s^h \qquad \text{(III.29)}$$

The division by 2 arises in accounting for species that are related by a 180° rotation about an axis coincident with the double helix axis, and which are therefore not distinguishable. However, n-h+1 of the complexes for a given n and h will not have staggered ends, and therefore will have C2 symmetry about the double helix axis. To account for these complexes, the molecular equilibrium constant should

be written

$$q = \sum_{h=1}^{n} \frac{(n-h+1)(n-h+2)}{2} \frac{\alpha}{V} s^h \qquad \text{(III.30)}$$

If the associated strands are antiparallel, the C2 axis is no longer coincident with the double helix axis, but rather is perpendicular to it, intersecting the complex midway along its length. For any h contiguous residues of one chain, if they are paired with h residues situated in the same relative position on the other chain, then the complex will have C2 symmetry and not be redundant. There will again be n-h+1 such complexes for a given h, and the molecular equilibrium constant is the same as the parallel case, equation (III.30).

Three Strand Association Models

With three strands associating, the expression for the system partition function analogous to equation (III.12) for two associating strands is

$$Q_{sys} = \sum_{i=0}^{Z/3} \frac{Q_{tri}{}^{i}}{i!} \frac{Q_{sin}{}^{(Z-3i)}}{(Z-3i)!} \qquad \text{(III.31)}$$

where Q_{tri} and Q_{sin} are the triplex and single strand partition functions. Solving for the maximal term of Q_{sys}, as in the two strand case, the expression for the molecular equilibrium constant q is

$$q \equiv \frac{Q_{tri}}{Q_{sin}^3} = \frac{i}{(Z-3i)^3} \qquad \text{(III.32)}$$

where i is the number of three strand complexes in the system.

Two-state triplex

In the case of infinite cooperativity between residues, the expression for the equilibrium constant becomes

$$K_{eq} = \frac{[complex]}{[free]^3} = \frac{\theta}{3[total]^2\ (1-\theta)^3} \qquad \text{(III.33)}$$

$$\theta = \frac{3[complex]}{[total]}$$

Where θ is again the fraction of total strands in complexes. Equation (III.33) can be solved for θ using numerical analytical methods (Salzer et al., 1958).

Zipper model for in-register triplexes

With three strands associating, they can all be parallel or one can be antiparallel to the other two. In either case, there will be no redundancy in complex enumeration if a molecular equilibrium constant like that for parallel in-register duplexes is used.

$$q = \sum_{h=1}^{n} (n-h+1) \frac{\alpha}{v^2} s^h \qquad \text{(III.34)}$$

Here $\alpha s/v^2$ is now the molecular association constant for three strands with the association of one residue from each strand. Solving for i in equation (III.32) using the algorithm in appendix II

$$i = \frac{Z}{3} \frac{-Z\left[\left\{1 + \sqrt{1 + \frac{4}{81Z^2q}}\right\}^{1/3} + \left\{1 - \sqrt{1 + \frac{4}{81Z^2q}}\right\}^{1/3}\right]}{3(6Z^2q)^{1/3}} \qquad \text{(III.35)}$$

Now the fraction of associated residues will be

$$\theta = \frac{3i\bar{h}}{Zn} = \frac{3i}{Zn} \frac{s}{q} \frac{\partial q}{\partial s} \qquad \text{(III.36)}$$

Substituting for i, in analogy to the duplex model

$$\theta = \left[\frac{1}{n} - \frac{\left\{1 + \sqrt{1 + \frac{4}{81Z^2q}}\right\}^{1/3} + \left\{1 - \sqrt{1 + \frac{4}{81Z^2q}}\right\}^{1/3}}{n(6Z^2q)^{1/3}}\right] \frac{s}{q} \frac{\partial q}{\partial s} \qquad \text{(III.37)}$$

and substituting for q yields

$$\theta = \frac{-s\partial J(s)/\partial s \left[\left\{1 + \sqrt{1 + \frac{4}{81\Omega}}\right\}^{1/3} + \left\{1 - \sqrt{1 + \frac{4}{81\Omega}}\right\}^{1/3} - (6\Omega)^{1/3}\right]}{(6\Omega)^{1/3} \, n \frac{\alpha}{v^2} J(s)}$$

(III.38)

where $\Omega = N_o^2 C_{tot}^2 \alpha$. Note that Ωs is the molar concentration equilibrium constant for association of three strands with association of one residue from each strand.

Relaxation of the in-register condition for triplexes

With three chains in the triplex, there can be no C2 axis perpendicular to the triple helix axis, so if one chain is antiparallel to the other two all of the triplexes are unique. Since h residues from each chain can be combined in n-h+1 ways, the molecular equilibrium constant will be

$$q = \sum_{h=1}^{n} (n-h+1)3 \frac{\alpha}{v^2} s^h \qquad \text{(III.39)}$$

In an ensemble of all possible triplexes with three parallel strands, those triplexes with their strands in register will have C3 symmetry about the triple helix axis. The other triplexes will be related to two other ensemble members by 120 and 240 degree rotations about the triple helix axis. The members of each such triplet are not distinguishable from one another in a quantum mechanical sense, and so the triplet should be counted only once in determining the molecular equilibrium constant for the system. If every ensemble element was a member of a triplet, this redundancy could be corrected for by dividing the equilibrium constant defined in equation (III.39) by the factor 3. This is analogous to division by 2 in the duplex model of equation (III.29). However, again in analogy to the duplex case, n-h+1 of the triplexes with h associated residue triples have in-register strands and C3 symmetry about the triple helix axis. They are not redundant. To account for them, 2(n-h+1) should be added to the coefficient for the term with h associated residue triples before division by 3. The molecular equilibrium constant becomes

$$q = \sum_{h=1}^{n} [(n-h+1)^3 + 2(n-h+1)] \frac{\alpha}{3V^2} s^h \qquad \text{(III.40)}$$

METHOD OF CALCULATION

Models were fit to the data using a simplex directed search technique (Nelder and Mead, 1965; Deming and Morgan, 1973). Briefly, the n unknown parameters (or factors) in the model for the transition (in this case, the transition enthalpies, entropies, and associated residue CD) are considered to define a factor space of dimension n. The simplex is a geometric figure defined by n+1 vertices in the factor space. Thus a simplex in two dimensions is a triangle. After establishing reasonable boundaries for the factor space, the simplex is placed at random within the boundaries to be searched. It then moves through the space according to strict rules governing which vertex should be discarded and how its replacement will be generated relative to the remaining hypersurface. The simplex functions by moving away from a vertex with a poor response. Nelder and Mead (1965) made the simplex more adaptable to the response surface by allowing it to expand or contract, depending upon results at the vertices. To judge a vertex, we calculated the CD predicted by the model for each experimental condition of temperature, oligomer concentration, and oligomer length utilizing the known experimental conditions and the factor values corresponding to the vertex coordinates. The predicted CD was compared to the observed value for each condition, and the sum of the squares (L2) or the absolute values (L1) of the differences comprised the vertex response. This allowed least squares fitting of a rather complicated nonlinear function of several variables. The method is easily adapted to some other function of the residual. The simplex search was typically repeated

in excess of 50 times to help determine the uniqueness of the solution, and to discriminate secondary local minima from global minima.

CHAPTER IV: THE CD OF HA OLIGOMERS

INTRODUCTION

As presented in Chapter II, the chiroptical transition of the HA polymer in aqueous-organic solvent at low pH as a function of temperature or solvent composition possesses considerable sigmoidal character. This is common in cooperative transitions of polymers, which typically involve large enthalpies and entropies as many residues essentially change conformation together as a single unit. The sharpness of cooperative transitions is due to the reduced energy required to propagate a conformational change through a polymer, relative to that needed for initiation of the change. Lack of symmetry observed about the midpoint of a melting curve, such as that in figure II.7, also argues for involvement of more than one energy in the transition process. Analyses of such transitions using a two-state model for each monomer can be expected to provide erroneous results, since cooperative systems will resemble two-state ones only under certain experimental conditions.

Suspected cooperativity in a polymer transition can be verified through studies of oligomers of the molecule, as has been done in the case of model polypeptides (Zimm et al., 1959; Nakamoto et al., 1974; Shaw and Schurr, 1975) and nucleic acids (Applequist and Damle, 1965; Martin et al., 1971; Nelson et al., 1981). We show here that the same approach can be used with polysaccharides. Studies of its oligomers in the acidic mixed-solvent system show that the transition of HA is a cooperative function of chain length. Moreover, the oligomer studies

reveal a sensitivity of the chiroptical transition to molecular concentration for many oligomers, showing that chain association is involved in the process. Associative transitions of long polymers can be insensitive to sample concentration, since even at infinite dilution the polymer effect will maintain a minimum concentration of the molecule in the vicinity of itself. This effective concentration and potential for intramolecular interaction will be governed by the length and inherent stiffness of the polymer chain, reflected in its persistence length and radius of gyration.

To pursue our analysis of cooperativity in the HA chiroptical transition, we have used an endo-β-hexosaminidase to cleave HA into a range of oligomers with the acetylated amino sugar at their reducing ends. Gel exclusion chromatography was then used to separate the species of various sizes. Acrylamide gel electrophoresis arose as the method of choice not only to determine chain lengths and polydispersity of HA in column fractions, but also to judge purity of sample materials regarding potential glycosaminoglycan contamination, and assess sample integrity during the course of our spectroscopic measurements.

Data describing chiroptical transitions of various HA oligomer species were analyzed in terms of two-state and cooperative transition models, the theoretical development of which was presented in chapter III. Our approach involved the use of a directed search algorithm known as a simplex (Nelder and Mead, 1965; Deming and Morgan, 1973) to determine those unknown parameters in equations III.11, III.26 and III.27 that minimized the fitting error of the model to the data. Although simplexing is a rather slow search technique, it is elegant

in its simplicity. Its primary advantage lies in its versatility of application. We use a single search algorithm here in conjunction with the rather complicated and diverse mathematical expressions of chapter III to locate fitting solutions for our various models. We are able to seek local as well as global solutions, and changing the error function helps us avoid erroneous results due to spurious data. From our analysis, we conclude that the chiroptical transition of HA in aqueous-organic solvent reflects a cooperative association of two strands, consistent with rheological (Welsh et al., 1980) and x-ray diffraction (Arnott et al., 1983) results.

MATERIALS AND METHODS

Hyaluronic acid isolated from streptococcus was purchased as the sodium salt from Calbiochem-Behring. Listed purity was greater than 99%, free of chondroitin, dermatan and keratan, and with protein contamination of less than 0.1%. The purity of this material is crucial, since it is being used as the substrate for a bovine hyaluronidase to cleave the polymer into oligomers. The hyaluronidase is a transglycosylase which is capable of degrading chondroitin as well as HA. The presence of contaminating chondroitin in the HA substrate will result in the creation of hybrid molecules.

Oligomers of hyaluronic acid were prepared essentially by the method of Cowman et al. (1981). The sodium salt of HA polymer was hydrated at 10 mg/ml concentration for two days at 4 degrees C in an aqueous solution 0.1 M in sodium acetate, 0.15 M in sodium chloride, pH 5.0. After warming to 37 degrees C, the polymer was digested with purified bovine testicular hyaluronidase (Worthington, hyaluronate 4-glycanohydrolase, E.C. 3.2.1.35) freshly dissolved in the same buffer at a concentration of 3000 IU/ml. The ratio of enzyme activity to HA polymer concentration in the digestion mixture was 10 IU/mg. Digestion was terminated after a period of two and one half hours by immersion in a boiling water bath for 15 minutes. The digest was then evaporated to dryness, solvated in 2 ml of 0.5 M pyridine-acetic acid solution, and chromatographed in the same solvent on a 950 ml packed bed volume of Bio Gel P60, -400 mesh (Bio-Rad). Fractions of 2.7 ml were collected at a flow rate of 3 ml/hr and screened for glucuronic acid content using the Bitter and Muir (1962) modifications of the

colorimetric assay of Dische (1947). Size heterogeneity of material in the fractions was assessed by polyacrylamide gel electrophoresis, and appropriate fractions were pooled, evaporated to dryness, solvated in distilled water, lyophilized, and stored at room temperature over phosphorous pentoxide.

Reagents for electrophoresis were purchased from Bio-Rad Laboratories, Richmond, CA. Tris, glycine and alcian blue dye were purchased from Sigma. Oligomers of hyaluronate were electrophoresed using a mini-slab gel apparatus (Idea Scientific) with 10 cm x 15 cm microscope slides and 0.8 mm spacers. Wells were 4 mm wide. The procedure used was essentially that due to Ornstein (Williams and Reisfeld, 1964). Briefly, a discontinuous system was used with chloride as the leading ion and glycine as the trailing ion. The stacking gel contained 3% acrylamide, 0.1% bis acrylamide in 125 mM Tris-HCl, pH 6.8. The separating gel was 20% acrylamide, 0.67% bis acrylamide in 375 mM Tris-HCl, pH 8.9. Reservoir buffer was 25 mM Tris-HCl and 190 mM glycine, pH 8.3. Sample buffer was that of the stacking gel, with addition of 1/10 the sample volume of 2 M sucrose with or without 0.2% bromophenol blue as tracking dye. Samples were electrophoresed at 50 V through the stacking gel, and then at 10-15 mA until the tracking dye was within 2 cm of the gel bottom, usually 4-5 hours. Gels were stained with 0.5% Alcian blue in water for 0.5-1 hour, and destained in water overnight. Gels sandwiched between glass plates were photographed and then scanned using a Zeineh soft laser scanning densitometer.

For measurements of temperature and concentration dependence of CD, hyaluronate oligomer samples were dissolved in water and then

dialyzed four times against a 250 to 500-fold excess of buffered mixed solvent at either 4 degrees C or at room temperature. All HA oligomer samples in acidic mixed solvent were buffered, containing 7.5 mM H_3PO_4 and 12.5 mM NaH_2PO_4. For repeated dialysis of sample volumes as small as 0.1 ml, the center of the cap of a 1.5 ml microcentrifuge tube was drilled out, and the tip of the tube was cut off. The cap served as a retainer ring to hold a piece of dialysis membrane (Spectra/Por 7, MW cut-off 2000) previously equilibrated in several changes of the dialysis solution across the top of the tube. The tube was then inverted and suspended inside the dialysis beaker with its top immersed in the solution. Sample oligomer was introduced onto the membrane through the open tip of the tube, using small bore teflon tubing attached to a syringe. Samples were introduced into and removed from spectroscopy cells using the same syringe after pre-rinsing with dialysate as appropriate. Circular dichroism spectra were recorded at a minimum of four different temperatures between 15 and 30 degrees C. Between sets of CD spectra, the sample concentration was assessed by optical density measurement using a Cary 15 or Cary 219 spectrophotometer with nitrogen purge. Sample concentrations were determined assuming an extinction coefficient of 9500 $M^{-1}cm^{-1}$ at 188 nm for the disaccharide of HA. After a set of CD spectra had been recorded, the sample material was recovered from the optical cell with dilution by an aliquot of the buffered mixed solvent, and was then redialyzed in preparation for the next series of measurements. This process was continued until measurements had been made for each oligomer at 10-15 different concentrations in the range of 0.2 to 20.0 mM in disaccharide. Following such a series of

measurements, the sample was recovered, dialyzed against distilled water, and lyophylized. This recovered material was then compared by electrophoresis to a portion of the original oligomer sample stock.

RESULTS AND DISCUSSION

Gel Electrophoresis of Oligomers

Figure IV.1 presents an example of the electrophoretic results obtained in this work, comparing oligomeric HA sample material recovered after CD spectroscopy (lanes 2, 7, 11) with an aliquot reserved from the original sample (lanes 1, 5, 10). The original sample was obtained by pooling fractions in a cut from a chromatographic separation of HA total digest. Lanes 3 and 9 in the figure contain total enzymatic digests of HA, with a range of oligomer sizes from six disaccharides (near the tracking dye front) to greater than fifty being resolved on the polyacrylamide gel. The assignment of an oligomer length to an electrophoresed band was made by comparing the electrophoretic profile of a chromatography fraction to its position in a plot of HA concentration versus fraction number. That plot presented well resolved peaks of concentration for oligomers shorter than ten disaccharides. The band assignments made by chromatographic peak counting were confirmed by electrophoresis of fractions using a protocol which resolves HA oligomers as small as seven disaccharides (Turner and Cowman, 1985).

In the example of figure IV.1, 70-80% of the material in the oligomer samples which is stained by alcian blue is 12 disaccharides long, with 10-15% contamination by each of the 11 and 13 disaccharide species. Change in these proportions is not significant when materials before and after CD spectroscopy are compared. There is relatively less of the 11 disaccharide contaminant present in material

Figure IV.1 HA oligomer electrophoresis. PAGE of HA oligomer samples comparing material from before (lane 1, 5, 10) with that recovered after a series of CD measurements (lanes 2, 7, 11). Neglecting any end effect on staining intensity, soft laser densitometer scans of the gel show 70-80% of the staining material to be 12 disaccharides in length, with 10-15% contamination by each of the 11 and 13 disaccharide species. These proportions did not change significantly during the course of the CD measurements. Similar results were obtained with other oligomers. Lanes 3 and 9 contained unfractionated digests of HA with testicular hyaluronidase. The fastest moving species clearly resolved behind the solvent front is six disaccharides long, as judged by peak counting in the chromatography profile and electrophoresis using a different protocol (Turner and Cowman, 1985).

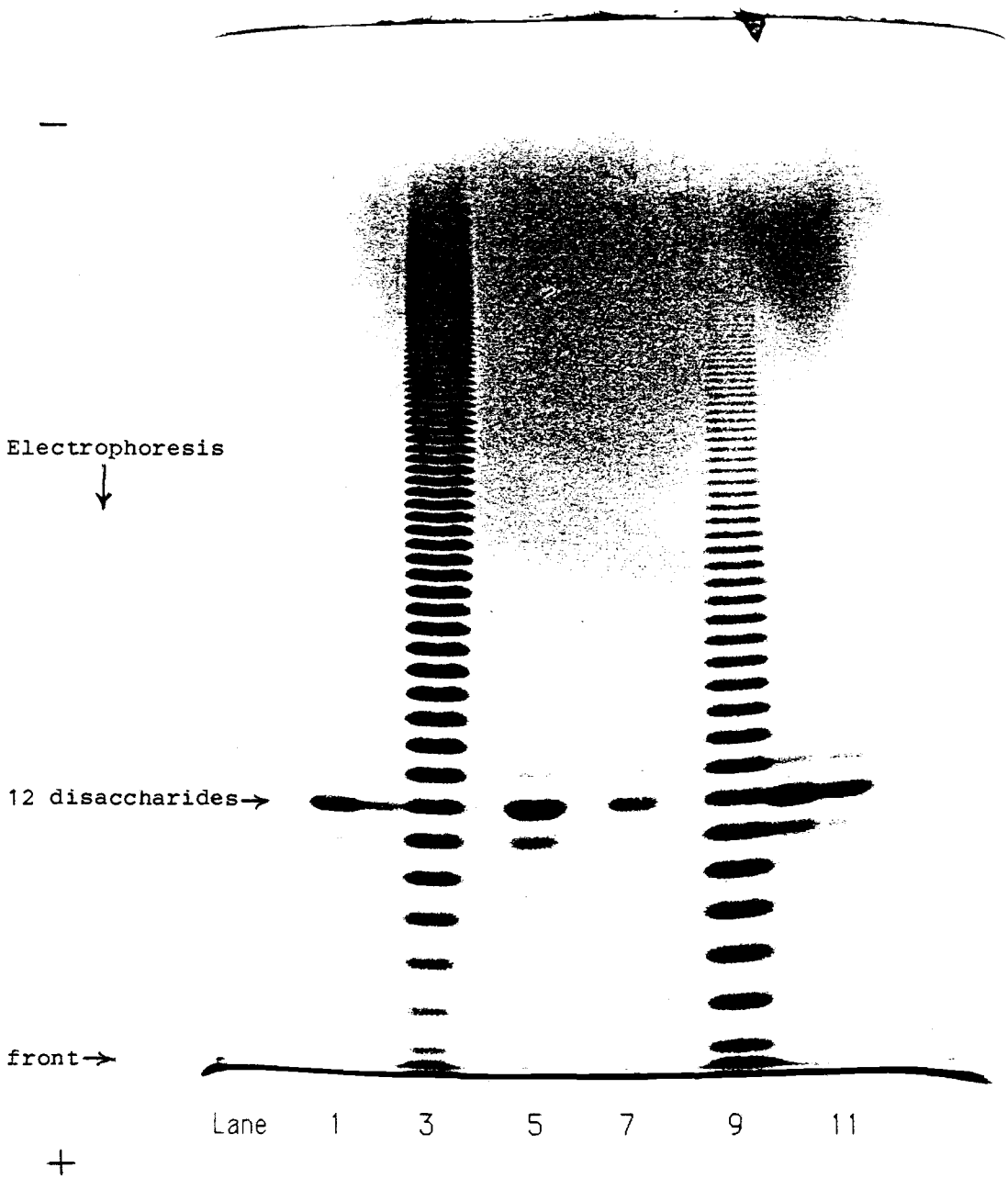

Figure IV.1

recovered after spectroscopy, due to the extensive dialysis of the sample performed between CD measurements (compare lanes 5 and 7 in figure IV.1). However, integration of densitometer scans of the gel show that the difference is not enough to affect the average HA oligomer length within the sample population.

As mentioned previously, lanes 3 and 9 in figure IV.1 contain total enzymatic digests of HA. Material in lane 3 was digested for 1.5 hours with bovine testicular hyaluronidase, and that in lane 9 was digested for 3.5 hours. The clarity of the oligomer ladders observed confirms the purity of the HA used as substrate in the digestion. The presence of contaminating chondroitin sulfate would be expected to result in the creation of hybrid oligomers due to transglycosylation events during digestion of the HA and chondroitin sulfate mixture. Chondroitin sulfate has a higher linear charge density than HA. Since charge density contributes to electrophoretic mobility, a hybrid oligomer would be expected to have a higher mobility than HA of the same chain length. In preliminary experiments involving digestion of mammalian HA and polyacrylamide gel electrophoresis of exclusion column fractions, two sets of stained bands were observed within the gel lanes. The set of bands with the higher mobility migrated out of phase with the bands of an unfractionated HA digest ladder used as a reference. It is likely that this set of bands is due to hybrid oligomers in the sample resulting from chondroitin sulfate contamination.

Chain length dependence of CD

The chain length dependence for the CD of HA in acidic mixed solvent is shown in figure IV.2, which presents the $\Delta\varepsilon_{190}$ for oligomers at 25 degrees C and approximately 5 mM concentration of disaccharide. The solid line in the figure is the CD predicted by a cooperative zipper duplex model. We will discuss this fit later. The dashed line in the figure represents the CD at 190 nm for various oligomers of HA in acidic aqueous solution at 25 degrees C. This CD is practically zero over most of the range of oligomer lengths investigated. The rise above baseline for oligomers from 12 to 16 disaccharides long does not exceed the noise level of the measurements. However, the negative CD for the smallest oligomers studied is clear and reproducible, approaching a value of $-1\ M^{-1}cm^{-1}$ for the five disaccharide species.

A similar dependence of HA CD on chain length has been noted previously in neutral aqueous solution by Cowman et al. (1981, 1983). Those authors examined spectra of NaHA oligomers in aqueous solution and found a CD dependence on chain length, which is most marked for those molecules smaller than 10 disaccharides.

Their observations indicated that an end effect was contributing to the total CD, according to the expression

$$\Delta\varepsilon_{obs} = \Delta\varepsilon_{int}(n-1)/n + \Delta\varepsilon_{end}(1/n) \qquad \text{IV.1}$$

where n is the number of disaccharides in the oligomer. Cowman et al. (1981) were able to show that oligomers of NaHA in aqueous solution do

Figure IV.2 Chain length dependence of HA CD. Chain length dependence of CD at 190 nm for oligomers of HA was recorded at 25 degrees C. Samples were dialyzed four times against a 200-fold excess of acidic ethanol-water solvent (45% by volume ethanol, 12.5 mM monobasic sodium phosphate, 7.5 mM phosphoric acid). Sample chain length was determined by electrophoresis (figure IV.1) and values are population averages, especially for longer oligomers. CD's for 9, 12, 14, 16, and 18 disaccharide species were interpolated at 5 mM concentration from plots of CD versus concentration for each oligomer (figure IV.3). CD was relatively insensitive to concentration for the other species plotted. The solid line represents the CD calculated from the best fit assuming a cooperative zipper transition with two associating strands which are allowed to slide relative to one another in the duplexes. Error bars represent noise levels of the CD measurements. The increase in negative CD observed for the shortest oligomers is likely due to an end effect such as that described by Cowman et al. (1981, 1983) in neutral aqueous solution.

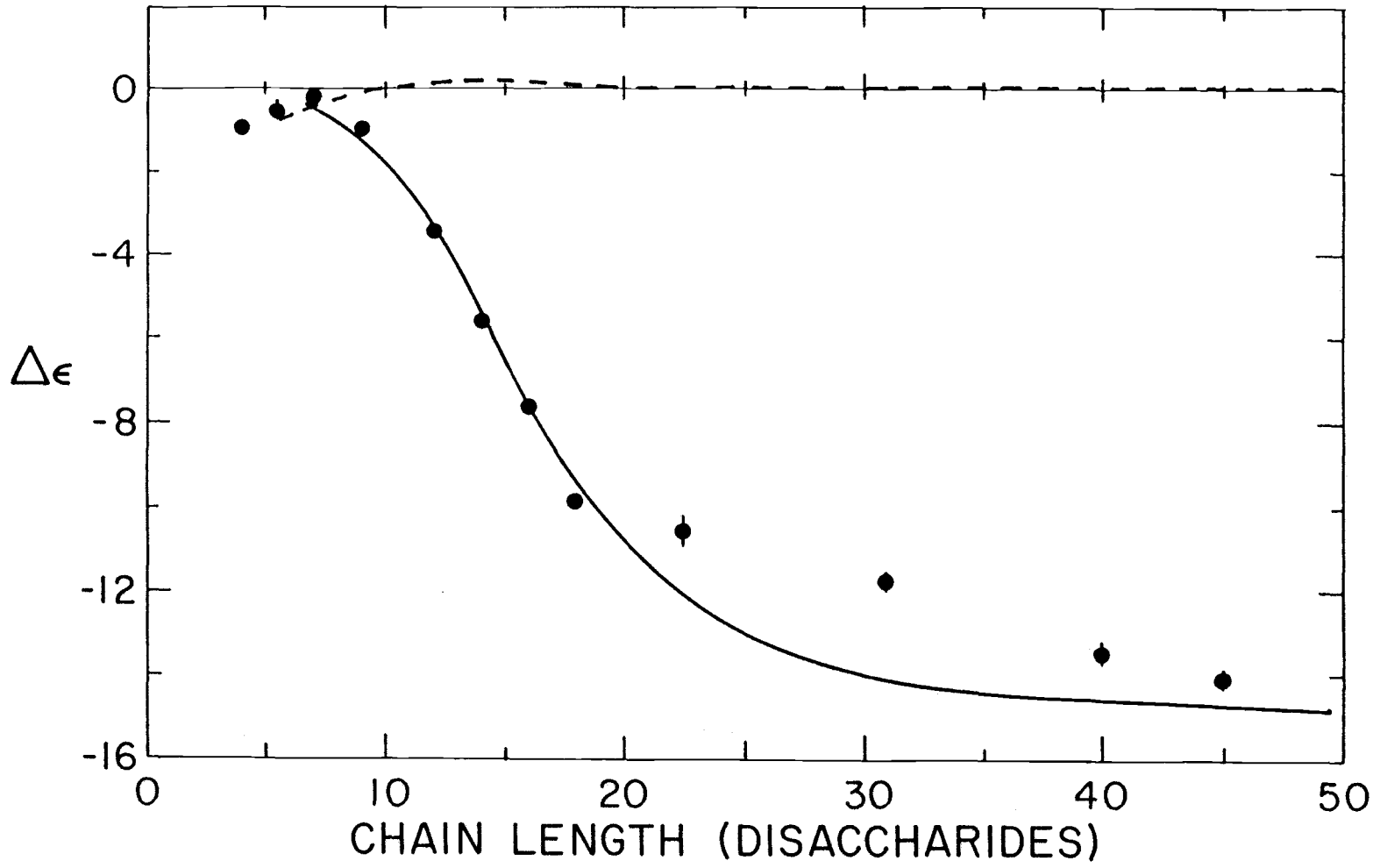

Figure IV.2

indeed display CD changes with length in the n-π^* transition region that are consistent with an end effect. We have observed a similar phenomenon in the π-π^* region of HA oligomer spectra recorded in acidic aqueous solution. As reflected in the dashed line of figure IV.2, the effect of chain length in acidic aqueous solution CD becomes apparent in oligomers smaller than 10 disaccharides. Assuming this CD to be due to an end effect as just discussed, the values of $\Delta\varepsilon_{190}$ for our smaller oligomers are less negative than those expected from calculations using the results of Cowman et al. (1983) for end residue contributions in the π-π^* transition region of HA spectra. Their values of $\Delta\varepsilon_{int} = -1.14$ $\underline{M}^{-1}cm^{-1}$ and $\Delta\varepsilon_{end} = -6.82$ $\underline{M}^{-1}cm^{-1}$ at 190 nm predict a $\Delta\varepsilon_{190}$ for five disaccharide HA of -2.28 $\underline{M}^{-1}cm^{-1}$, and for seven disaccharide HA, -1.95 $\underline{M}^{-1}cm^{-1}$. We observe values closer to -1.0 $\underline{M}^{-1}cm^{-1}$ and -0.6 $\underline{M}^{-1}cm^{-1}$, respectively. The difference is likely due to pH, since the measurements of Cowman et al. (1981, 1983) were performed in neutral solution, and our measurements are under acidic conditions. Buffington et al. (1977) have shown that CD in both the n-π^* and π-π^* transition regions for HA in aqueous solution becomes more positive as the hydrogen ion concentration increases. Therefore, the deviation of the dashed line in figure IV.2 from zero CD for short oligomers is likely due to the increased contribution of end residues relative to the internal residues of the HA chain, and not some type of secondary structural change.

Now let us consider the chain length dependence of HA CD in acidic aqueous-organic solvent, presented as the solid circles in figure IV.2. In oligomers shorter than five disaccharides, the 190 nm CD is significant and negative, becoming less negative with increasing chain

length until a minimum magnitude is reached at seven or eight disaccharides. With further increase in size of the oligomer the negative CD at 190 nm increases in magnitude, the rate of increase with chain length being maximal between 12 and 18 disaccharides. Again, the negative CD observed in oligomers six disaccharides and shorter is likely due to end residue contributions. The differences in CD between acidic aqueous (dashed line) and acidic mixed solvent (solid circles) in this size range are small but reproducible. They correspond to the type of perturbations observed in HA monomer spectra such as those of figure II.1 with introduction of organic solvent.

The large increase in negative magnitude for $\Delta\varepsilon_{190}$ with increasing chain length clearly arises from another source. It is more representative of the dependence on chain length expected for a conformational transition requiring cooperation between adjacent residues in the polymer. If this is true, we can use a theory describing polymer transitions such as that presented in chapter III to analyze this chiroptical change in HA as a function of chain length, temperature, solvent composition, or some other experimental variable.

One of the first steps in studying a transition observed in some physical characteristic involves defining the limits or endpoints of the characteristic at the extremes of the transitional variable. Such information is important for subsequent analysis. In figure IV.2 the transitional variable is HA chain length. At the extreme of short oligomers with seven disaccharides and less, the 190 nm CD of HA in aqueous-organic solvent at low pH (solid circles) is comparable to that in acidic aqueous solution (dashed line). Thus the CD of the

seven disaccharide species of HA appears to represent an (disordered oligomer) endpoint in the chiroptical transition of the molecule as a function of chain length at 25 degrees C and 5 mM concentration of the HA repeating disaccharide. The actual CD transition data which we wish to analyze include measurements of different size HA oligomers as a function of sample concentration. The disordered CD for a particular oligomer presumably is that observed at infinite dilution. In the case of the 7 disaccharide species, the CD is insensitive to oligomer concentration, as expected for an endpoint oligomer length. However, it is this insensitivity which limits the usefulness of short oligomers in investigating the transition. What we really need are the disordered CD endpoints, at infinite dilution, for longer oligomers of HA. These may not be practically attainable. We can approximate them with short oligomer spectra, but we already know that there is a CD end effect (represented by the dashed line in figure IV.2) which will introduce error. The identity of short oligomer CD in aqueous solution with that in aqueous-organic solvent suggests a better approximation. For a given oligomer of HA demonstrating concentration dependence of CD in mixed solvent, we shall use its acidic aqueous solution CD as the disordered structure endpoint at infinite dilution. We shall see that spectral data for 12 and 16 disaccharide HA show this to be a reasonable course.

The other limit of $\Delta\varepsilon_{190}$ for HA in buffered mixed solvent, that at infinite chain length, is not clearly indicated in figure IV.2. Measurements for oligomers averaging 40 and 45 disaccharides in length are -13.6 and -14.2 $M^{-1}cm^{-1}$, respectively. Measurements of HA polymer in the buffered mixed solvent might result in a larger value, but this

could not be verified. CD measurements of the polymer in acidic mixed solvent containing phosphate buffer were attempted, but the presence of linear dichroism (LD) signals in the samples made the results unreliable. Several approaches were used to attempt measurement of unoriented polymer gel which forms in the presence of counterions. These included squashing of the sample in a quartz window sandwich, melting the gel for introduction into cylindrical cells, and addition of an aliquot of concentrated buffer to deionized polymer solution within the cell. Samples were checked for orientation by measuring their LD. While it was not possible to obtain a sample that had no LD, the $\Delta\varepsilon_{190}$ of those samples with minimal LD signal was approximately -15 $\underline{M}^{-1}cm^{-1}$ at 25 degrees C. It should be noted that these samples were of deionized polymer solution in a cell, to which buffer solution was added. Diffusion of ions throughout the cell could not be verified. The substantial increase in negative CD intensity in the π-π^* region for HA in the buffered mixed solvent as opposed to unbuffered solvent suggests at least partial diffusion of counterions (data not shown). Since this endpoint is difficult to estimate, we will make it one of the unknown model parameters to be provided for us by the data fitting procedure.

Concentration dependence of oligomer CD

We wish to determine whether or not the conformational change in HA reflected by the chiroptical transition in mixed solvent involves strand association. In chapter II it was stated that the CD of HA polymer in acidic mixed solvent is insensitive to HA concentration.

However, examination of oligomers can reveal concentration effects not observable in longer molecules. In particular, we wish to study those oligomers lying midway through the chiroptical transition as a function of chain length over a reasonable concentration range for CD measurements. From figure IV.2, those oligomers are near 15 disaccharides in length, at 5 mM concentration of disaccharide.

In the size range of 10 to 16 disaccharides, oligomer CD spectra are quite sensitive to sample concentration under our experimental conditions. This sensitivity is displayed in figure IV.3 for the 16 disaccharide species at 30 degrees C. Analyses of such sets of spectra for orthogonal components, as done for HA polymer spectra with differing alcohol contents of solution, again detect only two significant components above the noise level of the measurements. The spectral series as a function of oligomer concentration in figure IV.3 is very reminiscent of the series presented in figure II.4 of HA polymer as a function of solvent composition, and presumably reflects the same conformational change. There are some differences between the sets of spectra. For example, the isodichroic point in the oligomer series near 207 nm, -2.5 $M^{-1}cm^{-1}$ is shifted relative to that of the polymer series with varying solvent composition (204.5 nm, -2.7 $M^{-1}cm^{-1}$). This difference is small but reproducible. It is likely due to combined perturbations of CD band energies and magnitudes by temperature and solvent composition differences between the two series. Figure II.2 shows that the effect of increasing temperature is to decrease the magnitude of negative CD in the n-π^* region of unordered HA. As seen in the GlcNAc monomer spectra (figure II.1b), addition of ethanol to the solution results in decreased CD magnitude

Figure IV.3 Concentration dependence of HA oligomer CD.

Concentration dependence of CD for the 16 disaccharide oligomer of HA was measured in acidic ethanol-water solvent at 30 degrees C. Sample concentrations on the figure in mM of disaccharide were determined using an extinction coefficient at 188 nm of 9500 $M^{-1}cm^{-1}$ for the disaccharide of HA, established by comparison of optical density and colorimetric data. This value did not change by more than 3% in acidic ethanol-water solvent when compared to acidic aqueous solution, and closely approximates the sum of the HA monomer coefficients under these conditions (9700 $M^{-1}cm^{-1}$).

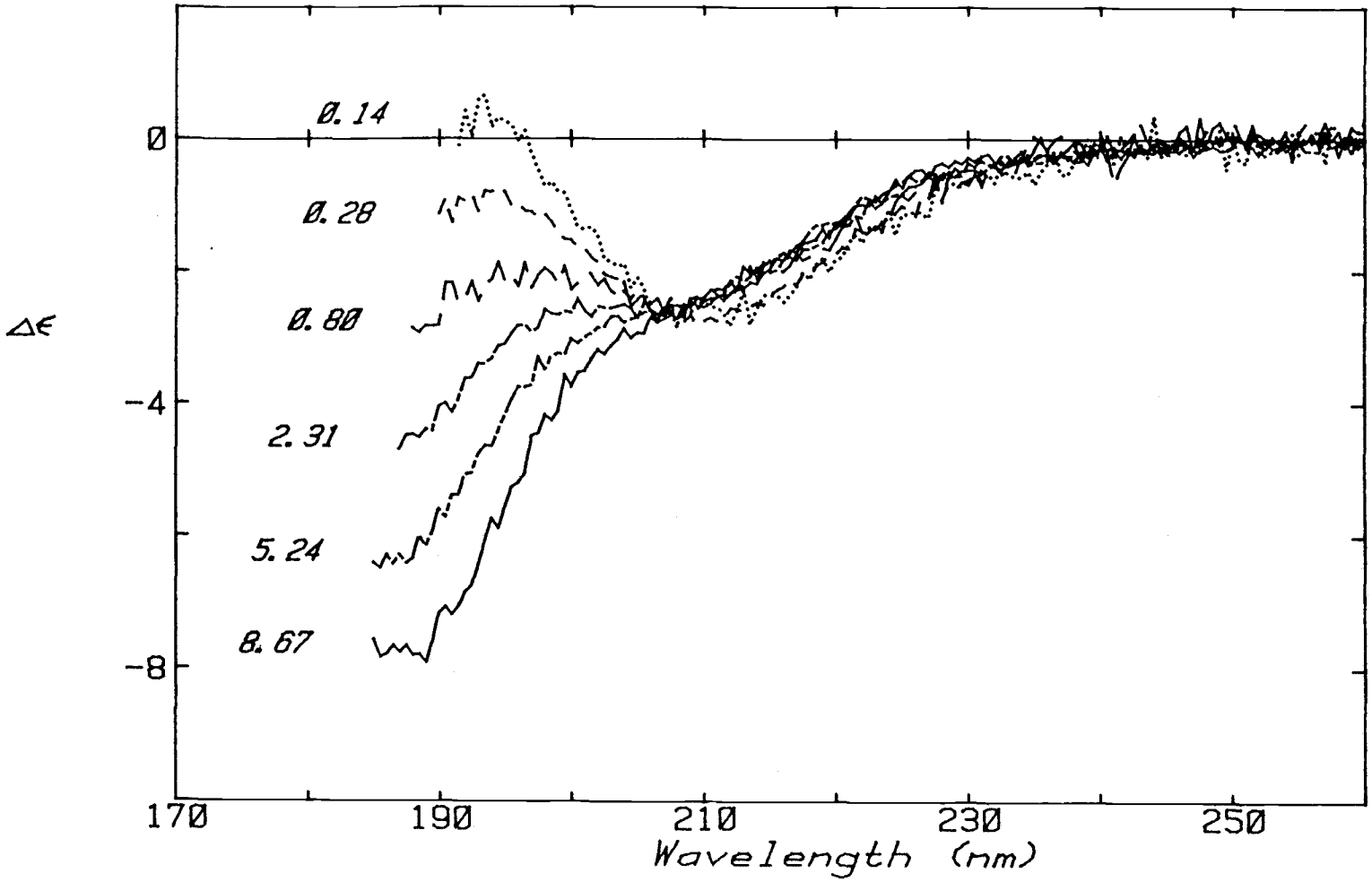
0.14
0.28
0.80
2.31
5.24
8.67
0
-4
-8
Δε
170
190
210
230
250
Wavelength (nm)

Figure IV.3

as well as a red shift of the band in the n-π^* region for GlcNAc. The result is a shift of the isodichroic point to longer wavelength and smaller magnitude in the oligomer spectral series, recorded at higher temperature and (average) organic solvent composition than the polymer series.

In order to analyze HA oligomer transition data, such as those in figure IV.3, in terms of transition models, it will be useful to estimate the endpoint spectra for the transition. A similar situation was encountered in chapter II, in which data describing the melting out of ordered structure in HA polymer in aqueous-organic solvent were analyzed using a two-state transition model. In that instance, the CD of HA in aqueous solution was used to represent the disordered polymer in aqueous-organic solvent. This was justified by the fact that spectra measured in acidic aqueous and acidic aqueous-organic solution at high temperature are essentially identical.

We shall again use aqueous solution CD spectra to represent disordered HA at a transition endpoint, but this time it is for oligomeric HA at infinite dilution. In figure IV.3 it is apparent that at low concentrations of 16 disaccharide HA, its CD in acidic mixed solvent begins to resemble that of HA in acidic aqueous solution. This resemblance is equally striking for the 12 disaccharide oligomer of HA. In figure IV.4, spectra are presented for 12 disaccharide HA in acidic aqueous solution and acidic mixed solvent at low oligomer concentration. The differences between the two spectra are barely larger than the error in the measurements. The chain association which results in conformational order in the acidic mixed solvent can be completely reversed upon dilution of the 12

Figure IV.4 CD of the 12 disaccharide oligomer of HA at high dilution. The CD spectrum of the 12 disaccharide oligomer of HA in acidic ethanol-water solvent was recorded at 15 degrees C and 0.214 mM concentration of disaccharide (----). For comparison, the CD spectrum of the oligomer in the absence of ethanol is also shown (——). Differences between the two spectra are small but reproducible. HA monomer (i.e. N-acetyl glucosamine and glucouronic acid) CD's show similar differences with and without ethanol present. Near identity of the spectra provides justification for using aqueous solution data to represent the high temperature or low oligomer concentration endpoint in the chiroptical transition.

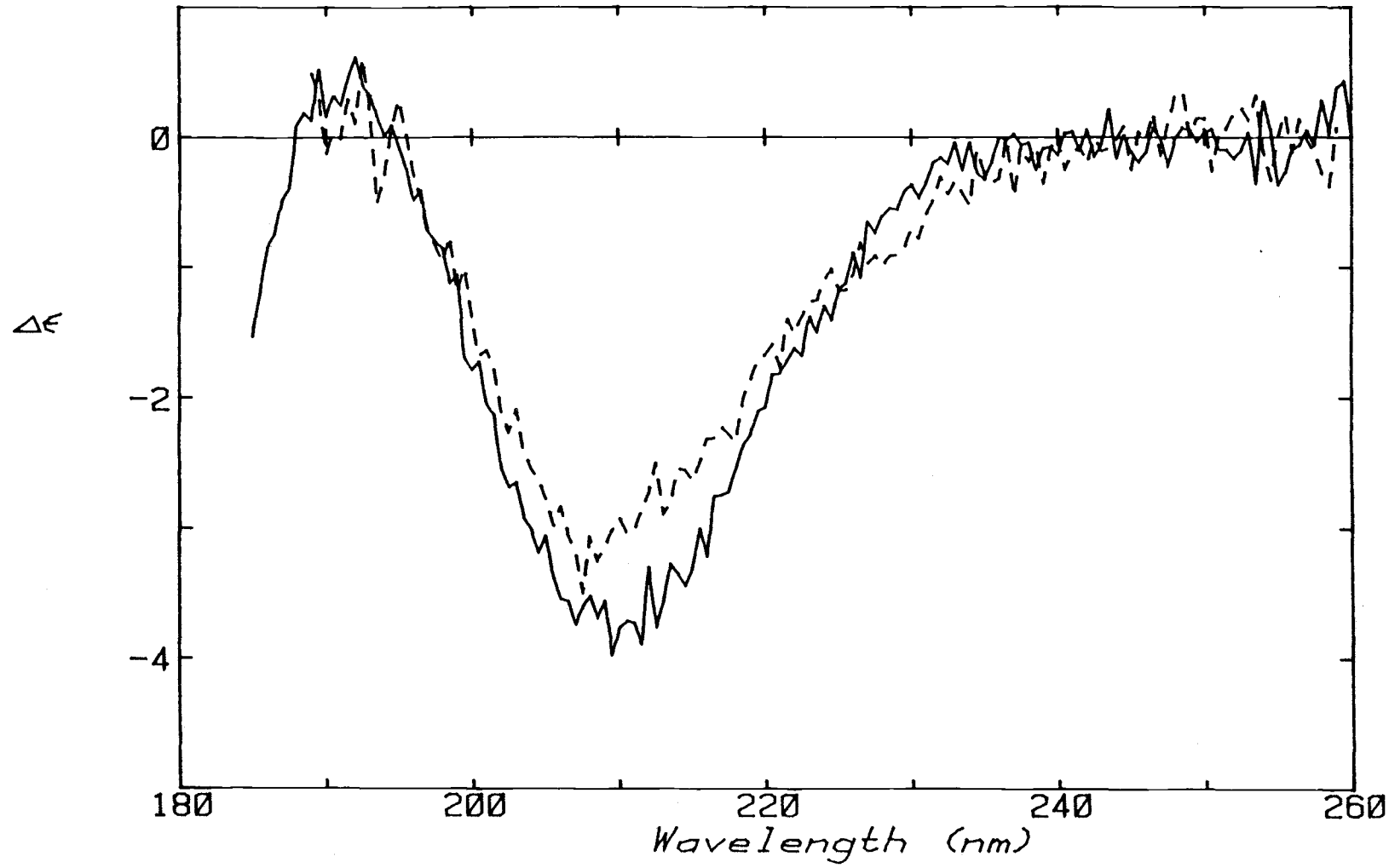

Figure IV.4

disaccharide HA. If we can use aqueous solution spectra to represent unordered HA oligomer in acidic mixed solvent, the task of fitting our data using the directed search technique is greatly simplified. We could simplify the search for solutions further by estimating the other transition endpoint, that of totally ordered HA oligomer. Such an estimate is more difficult to determine from the data than the infinite dilution endpoint, and we would have less confidence in any value(s) that we chose. Therefore we will let the ordered HA CD remain as an unknown to be determined by the data fitting procedure, along with the transition energies. We will see that cooperative model fits involving two strand association provide reasonable values for this unknown, while those values from three strand association models and simple two-state models are less so. Before proceeding to a discussion of the model fits, we will review our experimental methodology and see why the data for HA oligomer CD are not presented as a series of melting curves, as is typical in cooperative transition analysis. Rather, our data were recorded more as a series of concentration/dilution curves.

A common and simple method of analyzing cooperative polymer transitions involves measurements of melting curves for oligomers of the molecule at several different concentrations (assuming a strand association of some type). The initiation and growth energies can be determined from the slopes of the melting curves at the transition midpoints, and the shifts in melting midpoint temperatures as the oligomer concentrations are changed. Measurement of complete melting curves is not always feasible for the transition being investigated, however, and such is the case for HA in acidic mixed solvent. In

order to completely melt out the strong negative CD observed in HA in acidic mixed solvent, temperatures approaching 100 degrees C are required. At such elevated temperatures and low pH, there is a clear potential for ester formation involving the glucuronic acid residues in HA and the alcohol in the solvent. If ester formation should occur, it could be expected to affect the course of the conformational transition and its reversibility. This sensitivity to ester formation would likely increase with a decrease in chain length of the HA.

A second difficulty was noted in preliminary measurements of HA in acidic mixed solvent at high temperatures. During the course of a CD measurement, the solvent would evaporate from the sample solution and condense in the cell at the location of lowest temperature. This was typically in the neck of the cylindrical cell, away from the brass jacket. Not only would this affect the concentration of HA in the remaining sample solution, but the recondensed solvent likely has a different composition than that in the original sample, and so the (remaining) sample solvent would also have an altered composition.

Just as high temperatures are required to completely melt out ordered structure CD in long HA oligomers, so are very low temperatures required to completely order the shorter oligomers. Preliminary measurements showed that temperatures below -10 degrees C would be necessary to clearly establish the ordered structure CD spectrum for a HA oligomer as long as 14 disaccharides. At these temperatures there was evidence of sample aggregation with decreases in optical density and differential optical density, particularly at higher oligomer concentrations. In order to measure a series of CD

spectra to characterize the melting of an oligomer in the mixed solvent, a sample must be maintained over a wide range of temperatures for a minimum of many hours. This proved to be generally impossible, especially for measurements at higher HA oligomer concentrations since there was not sufficient sample available to completely fill the spectroscopy cells.

The temperature domain of our measurements is limited by the experimental conditions mentioned above. We may still be able to collect enough data for an analysis of the transition if we expand the concentration domain of our measurements. Thus HA oligomer spectra were recorded at many different concentrations over a modest range of temperature. This precluded potential difficulties with sample aggregation at very low temperatures required to attain an endpoint in the transition, as well as stability of the solvent at very high temperatures.

Use of a moderate temperature range also reduced the demand on the buffer during the course of the transition. The importance of carboxyl group protonation and its effect on the transition have been noted (Park and Chakrabarti, 1978c). In particular, these authors used titration of HA solutions, either with acid in mixed solvent, or with organic solvent component in acidic medium, to show that the conformational transition of HA promotes protonation of its carboxylate groups. The system therefore does not display normal acid-base behavior in the region of the transition. In the present work, if most of the carboxyl groups are either ionized or unionized in single strands of HA oligomer under the experimental conditions, and associated strands are sharing protons or totally protonated, then

melting or diluting out the ordered structure may result in proton absorption from or release into the medium. The resulting pH change could affect the further course of the transition. It is for this reason that all of the HA oligomer solutions are buffered with phosphate. This effect on solution pH is also minimized by measuring only part of the melting curve for a HA oligomer at any given concentration. At high sample concentrations, the effect of sample dilution on system pH is removed by the use of dialysis between sets of CD measurements. Upon repeating a series of measurements, there is no detectable inconsistency in HA oligomer spectra recorded as a function of concentration at a given temperature, whether the sample was previously dialyzed at 4 degrees C or at room temperature (20 degrees C). This is consistent with the presence of adequate buffering of the system.

Data fitting

The procedure used to fit data describing the transition is flexible in the choice of a particular model, while preserving the logic which determines the movement of the simplex through that model's vector space of unknown parameters. The basic design of the simplex directed search is quite simple, and in fact the most difficult aspect of its use in the present work lay in determining the starting position for the simplex in the vector space. To see why this is true, let us consider the example of fitting data with a cooperative two strand association model. We will be calculating the CD at simplex vertices as predicted by the model using equation

III.11. In equation III.11, we use aqueous solution CD to represent CD_{fr}, while CD_{com} is one of the vector space unknowns to be determined by the fitting. Equation III.26 is used to express θ for our choice of model. The unknowns in equation III.26 are α and s, which in turn are described by equations III.27 in terms of temperature, ΔH_{ind}, ΔS_{ind}, ΔH_{res} and ΔS_{res}. The enthalpies and entropies constitute the other four unknowns to be determined by the fit. Now, when we place the original simplex in the unknown vector space, we would like to choose the unknowns independently. However, we are limited by the computing power at our disposal, for the following reason. If energy terms such as ΔH_{ind} and ΔS_{ind} are placed at random within generous domains, it becomes likely that an exponential expression such as that for α in equations III.27 will either explode, or become vanishingly small. We avoid this situation by choosing the enthalpy and entropy as a unit. One of the two is placed at random within a generous domain, such as -50,000 to +50,000 cal/mole for ΔH_{ind}. The other unknown is randomly placed in a smaller domain, the size of which is determined so that an expression such as that for α in equations III.27 is calculable by the computer. In practice, the simplex procedure is attempted first without any linking of parameters in the initiation. As calculation errors are discovered, the algorithms for simplex initiation and calculation of fitting error are altered to eliminate difficulties with very large or very small intermediate values. The simplex does not waste time searching those regions of the vector space of unknowns which correspond to θ values of essentially 0 or 1 under the experimental conditions. It instead remains in those regions nearer to $\theta = 0.5$, where the data lie. If

the simplex is not placed at least partially within such a region, the error in fitting data by using model parameter values corresponding to each of its vertices is essentially constant. The search algorithm then has difficulty in determining appropriate movement for the simplex.

Fitting results

The results of simplex fitting our data with various transition models are presented in table IV.1. The solution listed for a given model corresponds to that with the smallest fitting error found in excess of 50 simplex trials, using the sum of either the least square (L2) or least absolute value (L1) of the residuals to judge the fit. The energy parameters with subscript ind are for the chain length independent or end effect contribution to the transition energy in two-state models, and the initiation step in cooperative models. The parameters with subscript res are the per residue contributions in two-state models, and growth step values in cooperative models. To be more precise, $\Delta H_{ind} - T\Delta S_{ind}$ in the cooperative models corresponds to $RT\ell n\alpha$ of Applequist and Damle (1963), and $\Delta H_{res} - T\Delta S_{res}$ corresponds to $RT\ell ns$. Since the equilibrium constant for strand association in the zipper model is αs, the enthalpy and entropy for this step are actually the sums of the listed initiation and growth values.

When we examine our results of fitting experimental data with various models, we can make the following observations:

1) Three strand association models provide smaller estimates of ordered structure CD than two strand models. This value is as small

Table IV.1 Transition Model Fits

	Energy Terms					
	length independent/initiation		per residue/growth		ordered	fitting
	ΔH_{ind} (cal/mole)	ΔS_{ind} (eu/mole)	ΔH_{res} (cal/moles)	ΔS_{res} (eu/mole)	CD_{com} ($\underline{M}^{-1}cm^{-1}$)	error
two-state models						
duplex	-9160	-32.8	-1590	-4.1	-12.2	14.1 (L2)
	-7290	-26.3	-1670	-4.4	-12.1	34.7 (L1)
	0.0	-1.82	-2249	-6.34	-12.1	14.4 (L2)
	0.0	0.0	-2162	-6.18	-12.4	18.7 (L2)
triplex	-580	-0.8	-4100	-11.5	-10.4	45.5 (L2)
cooperative zipper models						
duplexes						
parallel in-register	-14700	-53.5	-1320	-3.3	-14.5	13.0 (L2)
	-16300	-58.7	-1160	-2.8	-14.4	33.5 (L1)
antiparallel in-register	-14180	-51.5	-1345	-3.36	-14.3	13.0 (L2)
sliding	-18370	-66.8	-1088	-2.56	-16.4	12.8 (L2)
	0.0	-4.87	-2560	-7.50	-15.7	16.0 (L2)
	-19700	-71.3	-950	-2.1	-16.4	33.2 (L1)
triplexes						
parallel in-register	8335	28.9	-4936	-14.4	-11.0	40.7 (L2)
parallel sliding	8411	29.1	-5078	-15.0	-12.0	34.1 (L2)
	0.0	0.72	-4430	-12.8	-12.1	34.2 (L2)

as -10.4 $\underline{M}^{-1}cm^{-1}$ for the two-state triplex, and as large as -12.1 $\underline{M}^{-1}cm^{-1}$ in some cooperative triplex fits. The corresponding range for duplex models is -12.1 $\underline{M}^{-1}cm^{-1}$ to -16.4 $\underline{M}^{-1}cm^{-1}$.

2) Two-state models (infinite cooperativity) provide smaller estimates of ordered structure CD than the models with finite cooperativity.

3) The fitting error is higher in two-state models than in ones with finite cooperativity, and is higher in three strand models than in two strand models. Comparison of the least square (L2) errors in table IV.1 shows that the fitting error increases as the predicted magnitude of ordered structure CD decreases. The error magnitude is not simply a function of the ordered structure CD. If we examine the parallel in-register duplex model as an example, and fit the model by fixing the ordered CD at various values in the range of -13 $\underline{M}^{-1}cm^{-1}$ to -16 $\underline{M}^{-1}cm^{-1}$, we find that the the fits with smallest error are those with the ordered CD fixed near -14.5 $\underline{M}^{-1}cm^{-1}$.

4) In cooperative two strand models, the effect of allowing chains to slide within the complex is subtle. It involves a sharpening of the initiation step of the transition in the temperature domain, as the initiation enthalpy and entropy both increase. There is a concomitant broadening of the growth step as its enthalpy and entropy decrease. By contrast, the introduction of sliding chains in three strand cooperative models has considerably smaller influence on enthalpies and entropies.

5) Attempts to simplify models by removing any contribution of the initiation step to the transition enthalpy (i.e. $\Delta H_{ind} = 0.0$) consistently result in a larger fitting error as the total transition

enthalpy is absorbed by the growth step.

At first glance the fitting errors in table IV.1 may appear rather large. However, the total error presented is that resulting from fitting more than 150 experimental data points. The root mean square error per datum in the least square fits of the duplex models is therefore approximately 0.3 $\underline{M}^{-1}cm^{-1}$. The average error per datum in the least absolute value fits is even smaller. The similarity in model parameters resulting from the fit, whether the least square or least absolute value of the residual is being minimized, suggests that the data do not contain values grossly in error which are unduly influencing the solution.

Another way to see this, and achieve a more thorough understanding of the nature of the error, is by direct visual comparison of the theoretical model fits with the raw data. An example is presented in figure IV.5, which shows the fits of cooperative transition models with two or three strands associating parallel and in-register. The two strand model fit is represented by the solid curve. The choices of 12 disaccharide oligomer data at 30°C (panel a) and 16 disaccharide oligomer data at 15°C (panel b) show clear differences between the two models. In comparison of triple and double strand model fits, the triple strand fits were found inferior. Fits of triple strand models consistently had larger residual sums. Inspection showed these fits to be poor at the limits of the transition, i.e. for the shortest oligomer at the highest temperature used, and for the longest oligomer at the lowest temperature (see Figure IV.5). The value for associated residue CD predicted by these models was too small, being only -10 to -11 $\underline{M}^{-1}cm^{-1}$. Such values are exceeded in measurements of longer HA

Figure IV.5 Plot of 190 nm CD versus concentration for HA oligomers. The solid line is the least square fit of a cooperative transition model with two associating strands. The dashed line is the corresponding fit for three associating strands. Panel a is a plot of 12 disaccharide oligomer data at 30 degrees C, and panel b is that for the 16 disaccharide oligomer at 15 degrees C. Data form CD versus concentration were simultaneously fit for several oligomer lengths and temperatures, using a simplex algorithm to search the vector space of model parameters for the lowest sum of absolute value or square of the residuals. Fits of models with three associating strands, whether two-state or cooperative, consistently resulted in larger errors than the corresponding duplex models. Inspection revealed that the triple strand model fits were poorest for the shorter oligomers at higher temperatures (a), and the longer oligomers at lower temperatures (b). No such trend could be detected for the double strand model fits.

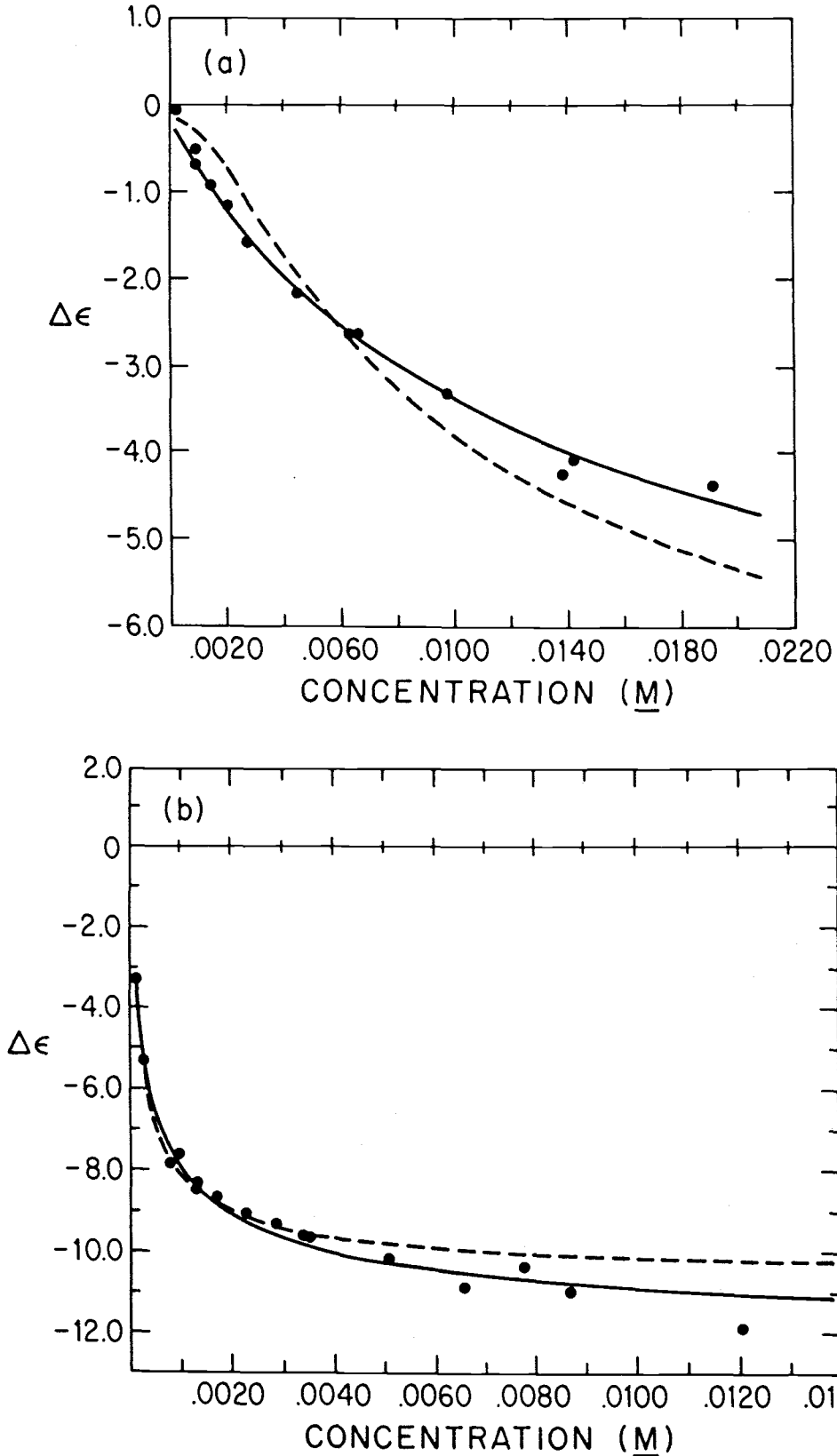
(a)
1.0
0
-1.0
-2.0
-3.0
-4.0
-5.0
-6.0
Δε
.0020
.0060
.0100
.0140
.0180
.0220
CONCENTRATION (M)
(b)
2.0
0
-2.0
-4.0
-6.0
-8.0
-10.0
-12.0
Δε
.0020
.0040
.0060
.0080
.0100
.0120
.0140
CONCENTRATION (M)

Figure IV.5

oligomers, which may well not be in the totally associated form and thus still not give the largest negative values.

In contrast to the comparison of double and triple strand models noted above, the effect of allowing chains to slide relative to one another in complexes, as opposed to remaining with chain ends in register, is rather subtle. The effect of allowing chains to slide relative to one another in the model is to shift the energy from the growth step to the initiation step of complex formation, with a concomitant increase in the estimated CD of the associated residues. Both sliding and in-register models provide reasonable fits by inspection, and their residual sums are comparable in magnitude. The one clear difference between the two is in their estimates of the CD for associated residues. A preference for the in-register model may be arrived at, due to measurements of fairly long oligomers of HA, on the order of 30-40 disaccharides, which reveal a CD at 190 nm between -14 and -15 $\underline{M}^{-1}cm^{-1}$. This CD value is in agreement with the in-register model. However, measurements of long HA molecules at lower temperatures are difficult for reasons already discussed, and we do not consider such results definitive.

Model parameter uncertainty.

The relatively small differences in model parameters observed among the various two strand association models in table IV.1 leads us to question their significance. We would like to know how sensitive our results are to the data, and the accompanying uncertainty in the parameters determined. How can we make this judgment?

One way to judge parameter uncertainty for a given model fit is by consideration of all of its simplex solutions. Remember that the results in table IV.1 are for the lowest error solution found in excess of 50 trials. When we inspect all of the solutions, we find in most cases that many solutions result with errors close to that of the smallest, and are located near that solution in the vector space of model parameters. Relatively few solutions further from this area with considerably larger error of fit also result, and occasionally a solution with very large error, usually driven against one of the limits establishing the region of the vector space to be investigated. We can estimate the uncertainty in our model parameters as that range of variation in them necessary to encompass all of those simplex solutions whose residual sums are within 0.3 of the smallest one. This value is arbitrary, but it was observed to include essentially all but those solutions with very large errors clearly lying along established vector space boundaries. This was true for searches minimizing either the least square or the least absolute value of the residual. The parameter uncertainties estimated in this way are ±3000 cal/mole for ΔH_{ind}, ±300 cal/mole for ΔH_{res}, ±15 eu/mole for ΔS_{ind}, ±1.3 eu/mole for ΔS_{res}, and ±0.5 $\underline{M}^{-1}cm^{-1}$ for CD_{com}. The difference between cooperative duplex models with chains that slide or remain in-register is at this limit of resolution.

As a second measure of our precision, we would like to know how variation in the data affects our results. Ideally, we would like to repeat our entire series of 157 experimental measurements for analytical comparison. This is impractical in terms of time and available sample material, and also unnecessary. After 157 separate

measurements, we have a very good idea about the uncertainties we experience in estimating sample temperature, concentration and CD. We can use this knowledge to perform mock repetitions of our measurements, as follows.

Given estimates of the variability in data such as temperature, CD measurement, and optical density (used for concentration), additional data sets are generated from the original employing Gaussian distribution functions centered at the data values. Data sets are also generated to represent systematic error in temperature, sample concentration, and oligomer chain length by appropriate alteration of the original data. These secondary data sets are then analyzed by the same algorithms used on the original set, and variation in solutions determined is noted. When we do this, we see that the solution for a particular model will be unchanged on average, but that the uncertainty in that solution's parameters is increased, perhaps by a factor of two. The cooperative sliding duplex model solution thus becomes indistinguishable from those for in-register duplex models, except in terms of the predicted value for CD_{com}. The cooperative duplex model solutions can still be considered distinct from the two-state duplex solutions, and the triplex solutions.

Two-state approximation of cooperative transitions.

Let us examine the usefulness of fitting cooperative transition data with two-state models (infinite cooperativity), and the limitations of such an approach. For purposes of discussion, a duplexing system of oligomers will be assumed. The analysis by

Applequist and Damle (1963,1965) of oligoriboadenylic acid transition data will be used as a reference point, one in which the two-state approximation is reasonable. We shall see that such an approximation is not suitable for analysis of the HA oligomer data, and that the reasonably small fitting error for the two-state model is spurious.

Applequist and Damle (1965) derived relations for analyzing cooperative transition data in that region of experimental parameters where the equilibrium constant for growth of the ordered structure, s, is much greater than one. A negative enthalpy and entropy for the growth process implies that this region will be found at lower temperatures. As s increases, it dominates the equilibrium constant for strand association (equation III.23), forcing the transition between single strands and complexes to lower strand concentrations. It may therefore be referred to as the transition region at high dilution. The importance of this region lies in the fact that the transition behaves in an all-or-none fashion. This is because s is large, which means that the associated free energy for formation of ordered residues in chain complexes is large and negative. Thus, once ordered structure of HA chains is initiated with chain association, essentially complete order within the complex is achieved. The expression for the fraction of ordered residues in the system is simplified considerably to

$$\theta = \frac{\left\{4C_{tot}N_o\alpha s^n + 1 - \sqrt{8C_{tot}N_o\alpha s^n + 1}\right\}}{4C_{tot}N_o\alpha s^n} \qquad \text{(IV.2)}$$

The transition midpoint $\theta = 1/2$ occurs when $C_{tot}N_o\alpha s^n = 1$, at which

point $\frac{d\theta}{d\ln s} = \frac{n}{6}$, and since $\ln s = \frac{-\Delta H_{gr}}{RT} + \frac{\Delta S_{gr}}{R}$,

$$\frac{d\theta}{dT} = \frac{d\theta}{d\ln s}\frac{d\ln s}{dT} = \frac{n\,\Delta H_{gr}}{6RT^2} \qquad \text{(IV.3)}$$

where T is the temperature of the transition midpoint. Equation IV.3 and other relations for the transition at high dilution have been successfully used in analyses of nucleic acid oligomer transitions (Nelson et al., 1981, Morden et al., 1983, Martin et al., 1971). However, it is difficult a priori to determine the experimental conditions of temperature and sample concentration under which the foregoing simplifications are valid. They may not be practically attainable for our system.

Consider the present case for example. In figure IV.6 are presented population distributions calculated for the HA sliding zipper dupleully used in analyses of nucleic acid oligomer transitions (Nelson et al., 1981, Morden et al., 1983, Martin et al., 1971). However, it is difficult a priori to determine the experimental conditions of temperature and sample concentration undend d correspond to values of s encountered by Applequist and Damle in the transition of the 11 residue oligomer of adenylic acid between 324 degrees K (s = 7.9) and 341 degrees K (s = 4.0). The distributions presented in panels b, c, and d of figure IV.6 are insensitive to chain length between 10 and 20 units, in the sense that longer oligomers simply have the population fractions shifted to higher n corresponding to the increase in oligomer length. Therefore, the fraction of 16 residue complexes with 13 ordered residues per complex

Figure IV.6 Duplex population distribution. The fractions of duplexes formed by 16 disaccharide oligomers of HA with various numbers of ordered disaccharides were calculated using the cooperative sliding duplex model and assuming $\Delta H_{init} = -18370$, $\Delta S_{init} = -66.8$, $\Delta H_{gr} = -1088$ and $\Delta S_{gr} = -2.56$. Panels b-d span the range of the growth reaction equilibrium constant s encountered in the melt of an 11 residue oligomer duplex of adenylic acid, analyzed by Applequist and Damle (1965).

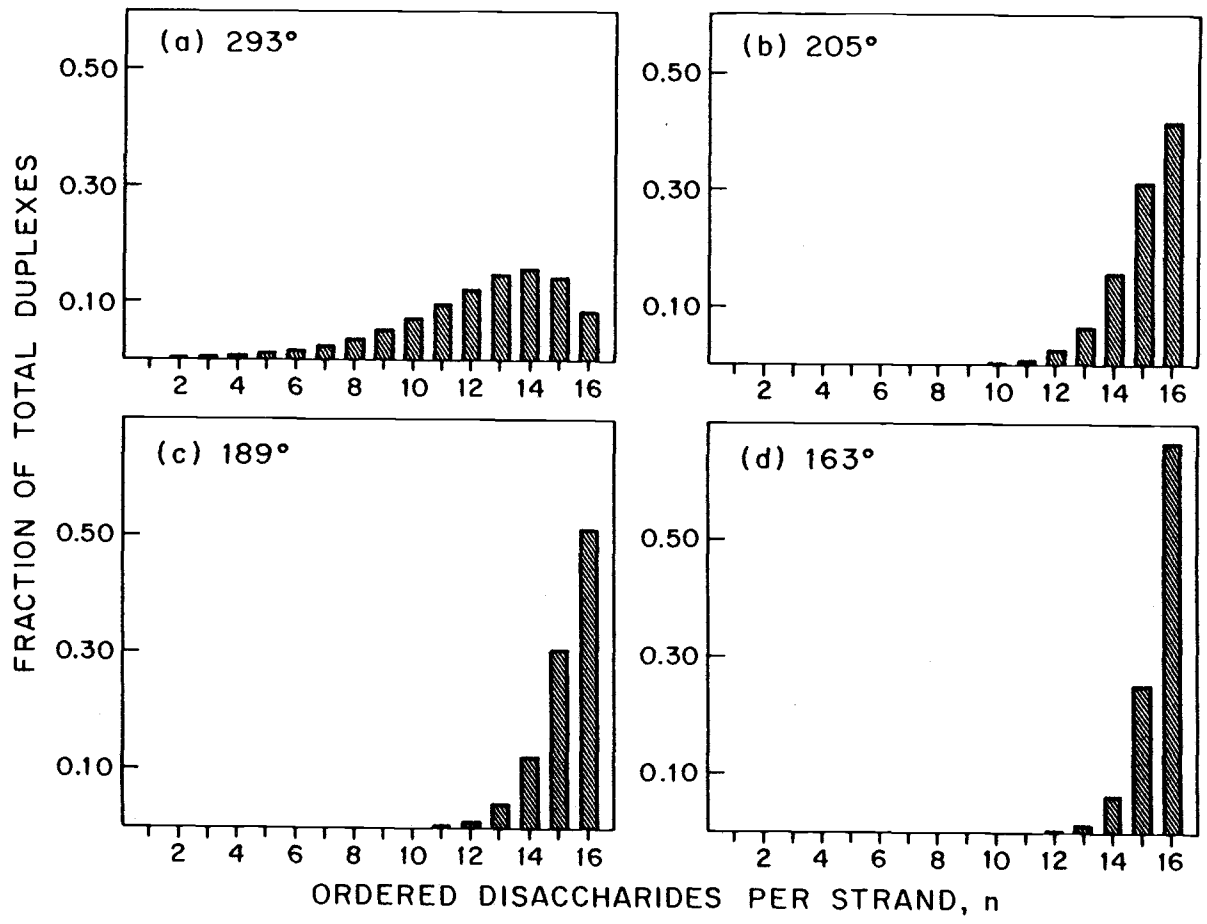
(a) 293°
(b) 205°
(c) 189°
(d) 163°
0.50
0.30
0.10
2 4 6 8 10 12 14 16
FRACTION OF TOTAL DUPLEXES
ORDERED DISACCHARIDES PER STRAND, n

Figure IV.6

is essentially the same as that for 11 residue complexes with 8 ordered residues per complex.

This is less true under conditions as in panel a of figure IV.6, at which s = 1.74 and there are an appreciable number of complexes present with few ordered residues. Applequist and Damle used their calculated population distributions to show that essentially all of the residues were associated or ordered in the multi-strand complexes, and that therefore the two-state approximation was a reasonable one. To achieve similar population distributions for oligomers of HA in the aqueous-organic solvent system at low pH, experimental temperatures in the range of 163 to 205 degrees K are necessary. The data for HA oligomers were recorded in the temperature domain of 288 to 303 degrees K, corresponding to duplex population distributions similar to that of figure IV.6a. The presence of duplexes with one half or fewer of their repeating units ordered suggests that a simple two-state analysis will not prove adequate.

This does not mean that a two-state model fit of the data will necessarily have a high error. In fact, in Table IV.1 the least square error in fitting a two-state duplex transitional model is seen to be comparable to the error in fitting the cooperative duplex models. The major reason that the two-state fit does not have more error is due to the rather narrow temperature range used. The two-state model does provide an erroneous result when applied to cooperative transition data that is not at high sample dilution, however. The manner in which it does so can be appreciated by considering the degree of residue association in the system duplexes, summarized by the average number of ordered residues in a complex, $\bar{h}$.

The $\bar{h}$ will not change very much through the temperature range of a cooperative transition when the associating strands are very dilute and the temperature is low so that s >> 1. It is for this reason that the transition appears to be a two-state one. For the oligoadenylic acid 11 mer data analyzed by Applequist and Damle, $\bar{h}$ changes approximately 10% from 273 to 343 degrees K. In contrast, the change in $\bar{h}$ for the 12 disaccharide oligomer of HA in acidic mixed solvent is more than two times as large over a comparable change in temperature. However, CD data were collected only between 288 and 303 degrees, in which range the change in $\bar{h}$ is approximately 5%. Since the CD of the average duplex with $\bar{h}$ ordered disaccharide pairs does not change very much in the region of temperature explored here, it may be expected that a two-state model would fit the data quite well. The CD endpoints for such a fit would correspond to the separated strands, and complexes with $\bar{h}$ ordered disaccharide pairs.

Difficulty in fitting data with a two-state model arises when several oligomers of different lengths are considered together. If the data are recorded at temperatures close to that at which s equals 1, then $\bar{h}$ can be expected to change appreciably with chain length. In the case of HA oligomer data, $\bar{h}$ decreases from .73 to .61 as the chain length is decreased from 16 to 9 disaccharides, at 303 degrees K. The result is an apparent decrease in the ordered (associated) residue CD with oligomer length, as determined by fitting the data with a two-state model. Without introduction of an end effect to account for the changing limit of associated strand CD, it should become difficult to fit all of the oligomer data with a single presumably average value. Close examination of two-state fits to the 190 nm CD data

shows that the errors are often small but also apparently not random. A second expectation of fitting with a two-state model is that since it is assumed all of the residues are ordered in the complexes, the predicted CD for associated residues will be too small in magnitude. It can be seen in Table IV.1 that the ordered residue 190 nm CD predicted by the two-state model is appreciably smaller than that predicted by the cooperative duplex models. In fact, the two-state fit value is approximately 75% of the cooperative fit value, corresponding to the average of .72 for $\bar{h}$ among duplexes of the oligomer lengths and temperatures encountered in the data.

Extrapolation to polymeric HA.

Figure IV.2 presents the chain length dependence of HA CD at 190 nm, compared with the CD predicted using the sliding duplex zipper model with thermodynamic parameters from Table IV.1. The model has difficulty in predicting the CD of oligomers longer than approximately 20 disaccharides. The CD results for HA samples averaging 40 and 45 disaccharides in length are -13.6 and -14.2 $\underline{M}^{-1}cm^{-1}$ respectively, which are more in agreement with the prediction of theory than the CD's for 22 and 31 disaccharide species in figure IV.2. It is not clear which results are more reliable, but all of the experimental values are smaller in magnitude than the predictions. Potential reasons for the discrepancy include applicability of the zipper model and existence of associated species higher than duplexes.

The zipper model as presented by Applequist and Damle disallows regions of unordered residues surrounded by ordered regions in the

polymer. As chain length increases in an oligomeric series, the likelihood of such a stretch of unordered residues will increase until in the infinite polymer such regions will alternate with ordered ones, the average lengths depending upon the equilibrium constants for initiation and growth of ordered regions. This behavior accounts for the failure to observe infinitely sharp transitions in long polymers.

In longer oligomers of HA, it is possible that the zipper model used to analyze the data is no longer adequate to represent the states accessible to the molecules in the system partition function. Therefore, we also analyzed the data set using a matrix-generated partition function of the type originally due to Zimm and Bragg (1959) and used to analyze transition data for collagen model peptides (Shaw and Schurr, 1975). This model allows "bubbles" of unordered residues within ordered regions. Unordered residues in the bubbles were statistically weighted the same as those at the chain ends (i.e. unity) in the absence of knowledge about loop probabilities. A third energy must be introduced when considering molecular states allowing for bubbles, since at least the entropic change associated with initial strand pairing can be expected to be different from that change when an ordered residue pair is formed defining an open loop in two strands which are already associated.

When we fit our experimental data using this matrix-generated partition function, we find that the thermodynamic parameters for growth and initiation of ordered regions are not very different for the bubble versus the zipper model. The enthalpy for complex initiation is somewhat larger in the bubble model, being -21200 cal/mole as opposed to -18400 cal/mole for the sliding zipper duplex

model. The entropy for complex initiation is essentially unchanged, at -67.4 eu/mole versus -66.8 eu/mole. Growth enthalpies for the two models are essentially the same at -1130 and -1088 cal/mole, while the bubble model growth entropy is a bit larger than that for the zipper model, at -3.04 versus -2.56 eu/mole. As a result, the bubble model predicts disorder of residues within strand complexes at a lower temperature than does the zipper model, but with complete dissociation of the strands occurring at a higher temperature or lower strand concentration.

In contrast, the bubble model predicts a considerably larger value for CD_{com} than the zipper models, at -22.2 $\underline{M}^{-1}cm^{-1}$. As the result, the fit of the bubble model to HA oligomer data as a function of chain length as in figure IV.2 is virtually identical to that of the zipper model. The CD predicted for longer oligomers is larger than that experimentally observed.

A second potential explanation for the discrepancy between observed CD in long oligomers, and that predicted by the model fits of shorter oligomer CD, is the possible existence of HA oligomer complexes containing more than two strands. Once we have estimated the energy parameters for the transition with the zipper model, we can use expressions similar to that in equation 24 of Applequist and Damle (1965) to estimate the equilibrium constants for formation of multi-stranded species. When we do this, we find for example that in excess of 30% of our sample molecules may participate in three-strand complex formation under our experimental conditions, as an oligomer forms a double-stranded region with more than one partner. The analysis becomes complex, and predicting the CD associated with such

structures is difficult. It is clear that the possibility exists for less complete pairing of HA strands because of this circumstance, especially for longer oligomers, with a concomitant decrease in the contribution of ordered structure CD to the total signal. It would be preferable to avoid this difficulty by investigating the transition in that region of sample concentration and temperature where it behaves in an all-or-none fashion. However, we have already seen that the two-state limit region for the present system is inaccessible, due to the extremely low sample concentrations and temperatures required. In any event, we cannot identify the region in which a cooperative transition behaves as two-state, or the region in which multi-strand complexes become a complicating factor, until we have estimated the transition parameters as in the present work.

CONCLUSIONS

Let us review our experimental results by listing the conclusions we can draw from them regarding the behavior of HA in aqueous-organic solvent at low pH.

1. A structural transition in HA takes place as the dielectric constant of the medium is reduced. We conclude this from the change in HA CD with increasing organic component of the solvent, depicted in figure II.6. The change in structure reflected in the CD is not localized within individual residues but involves the polymer as a whole, since the CD of HA monomers is only slightly affected by the presence of organic solvent (figure II.1) and cannot explain the polymer result.

2. HA is more ordered in aqueous-organic solvent than it is in aqueous solution. We conclude this from the fact that addition of heat energy to an aqueous-organic solution of HA results in a CD spectrum resembling that of HA in aqueous solution (figure II.10). We expect this energy to disrupt or denature ordered structure in the polymer, and as we add heat we arrive at essentially the spectrum found for HA in aqueous solution at the same high temperature. Further evidence for the relative disorder of HA in aqueous solution lies in the rather small, gradual change in its CD as a function of temperature (figure II.5).

3. During the course of the structural transition in aqueous-organic solvent, residues of HA can exist in one of two different states. The presence of isodichroic points in series of CD spectra describing the transition (figure II.6, figure IV.3) suggests

only two environments for HA chromophores. Each of these corresponds to a different local geometry. We can confirm the existence of only two chiroptical states by analyzing the CD spectra for orthogonal components. When we do this, we find only two pieces of information in a series of spectra, as described in figure II.8 for the transition of HA as a function of solvent composition. This result is important for our development of models to fit the experimental data.

4. The transition of HA is a cooperative one. The sharpness of the CD change observed upon melting the ordered structure of the polymer (figure II.11) indicates that the transition enthalpy and entropy are quite large. The resulting transition energy exhibits a large magnitude change over a modest temperature range. We definitively show that many residues of HA must cooperate to form the ordered structure with large transition enthalpy and entropy, by presenting the effect of chain length on the CD (figure IV.2). The large increase in negative CD which reflects assumption of ordered structure in aqueous-organic solvent first begins in oligomers of 8 or 9 disaccharides and quickly increases thereafter with oligomer length. It clearly does not result from a simple end effect.

5. The assumption of ordered structure involves strand association. This is clear from the concentration dependence of HA oligomer CD in aqueous-organic solvent (figure IV.3). For the 12 disaccharide species of HA, we can completely dilute out the CD of the ordered structure for the molecule in mixed solvent, and arrive at the spectrum observed in aqueous solution (figure IV.4). We do not observe such an effect in the polymer because our ability to reduce the effective sample concentration is limited. It is true that we can

reduce the HA strand concentration as much as we wish, but the polymer effect establishes a lower limit to the local residue concentration as the molecule hairpins or loops upon itself.

6. The association reaction involves two strands of HA. We conclude this from our results of fitting transition data with various models. The two strand association models consistently fit the data with smaller error then their three strand counterparts. Two strand models predict larger negative values for the CD of the ordered structure at 190 nm, consistent with experimental measurements. In plots of $\Delta\varepsilon_{190}$ versus concentration for shorter HA oligomers the three strand models predict a region where the second derivative is negative, which is not displayed by the data (figure IV.5a). In fact, none of the measured curves of $\Delta\varepsilon_{190}$ versus concentration is fit better by three strand models than by two strand ones. The reverse is consistently true.

7. The ordered structure in mixed solvent has a $\Delta\varepsilon_{190}$ estimated at -16.4 ± 0.5 $\underline{M}^{-1}cm^{-1}$. The enthalpy and entropy for growth of ordered structure are approximately -1000 ± 300 cal/mole and -2.25 ± 1.3 eu/mole, respectively. The corresponding values for initiation of order are -20000 ± 3000 cal/mole and -71 ± 15 eu/mole. These estimates result from fitting experimental data with a cooperative two strand zipper model which allows relative sliding of chains in duplexes. It has the smallest fitting error of all the models tested.

8. The ordered structure does not require the presence of metal cations. This follows from the appearance of ordered structure CD for deionized HA polymer (figure II.6). The result is important for the comparison of experimental conditions under which we made our CD

measurements with those under which fibers of HA were drawn for x-ray diffraction measurements.

Only one fiber diffraction pattern is presently proposed to represent a double-helical structure in HA. The fiber from which it was obtained was drawn at low pH in the presence of various monovalent cations (Sheehan et al., 1977). A recent reanalysis of the data suggests that the cations lie between double helices in the fiber, and that carboxyl groups are protonated to the extent that they can provide hydrogen bonding links between paired strands within double helices (Arnott et al., 1983). Our results are consistent with this model. Cations are not required for the appearance of ordered structure CD. If they are present, HA polymer will form a soft gel in the aqueous-organic solvent. This may be due to higher order association of double-helical domains to create a network, as suggested by Rees and co-workers (Robinson et al., 1980; Rees and Welsh, 1977) to explain features of the self-interaction in carrageenan solutions.

Rheological studies demonstrating network formation in HA solutions are also consistent with our results. Welsh et al. (1980) showed that network character of HA which is accentuated by lowering pH or reducing water content of the solution can be specifically inhibited by oligomers of the same copolymeric sequence. Oligomers 2-4 disaccharides long did not affect network formation, while those 60 disaccharides in length inhibited or abolished it. This was presumably due to the long oligomers competing with and breaking polymer-polymer junction zones, which small oligomers could not do. We would predict from our CD results that the critical size for

competition of oligomer with polymer in these junctions (formed in our solution conditions) is the size that can first form ordered structure (in the range of 11-15 disaccharides). This length may be increased at lower alcohol content of the solvent.

The ability of HA to self-associate, as well as interact with other glycosaminoglycans (Turley and Roth, 1980) and proteins (Hascall and Heinegard, 1974; Turley, 1982; Turley et al., 1985) may be critical to its physiological function in processes such as cell-substrate adhesion, cell motility, and tissue matrix organization. The existence of double-stranded HA under physiological conditions is speculative at present. The small fraction of total polymer residues necessary to create a transient network through junction zones may escape optical detection at neutral pH in aqueous solution. Rheological evidence of its existence is still observed (Morris et al., 1980).

BIBLIOGRAPHY

Anderson, N.S., Campbell, J.W., Harding, M.M., Rees, D.A. and Samuel, J.W.B. (1969) J. Mol. Biol. 45, 85-99.

Applequist, J. and Damle, V. (1963) J. Chem. Phys. 33, 2719-2721.

Applequist, J. and Damle, V. (1965) J. Am. Chem. Soc. 87, 1450-1458.

Arnott, S., Scott, W.E., McNab, C.G.A. and Rees, D.A. (1974) J. Mol. Biol. 90, 253-268.

Arnott, S., Mitra, A.K. and Raghunathan, S. (1983) J. Mol. Biol. 169, 861-872.

Atkins, E.D.T., Phelps, C.F. and Sheehan, J.K. (1972) Biochem. J. 128, 1255-1263.

Atkins, E.D.T., Brown, J.H., Landsall, J.M., Nieduszynski, I.A. and Sheehan, J.K. (1973) J. Polymer Sci. part C, 42, 1513-1520.

Atkins, E.D.T., Isaac, D.H., Nieduszynski, I.A., Phelps, C.F. and Sheenan, J.K. (1974) Polymer 15, 263-271.

Atkins, E.D.T., Meader, D. and Scott, J.E. (1980) Int. J. Biol. Macromol. 2, 318-319.

Balazs, E.A. (1958) Fed. Proc., Fed. Amer. Soc. Exp. Biol. 17, 1086-1093.

Balazs, E.A. (1966) Fed. Proc. Fed. Amer. Soc. Exp. Biol. 25, 1817-1822.

Balazs, E.A. and Gibbs, D.A. (1970) In Chemistry and Molecular Biology of the Intercellular Matrix (Balazs, E.A., ed.), vol. 3, 1241-1253, Academic Press, N.Y.

Barrett, T.W. (1975) Biochim. Biophys. Acta 385, 157-161.

Barrett, T.W. and Harrington, R.E. (1977) Biopolymers 16, 2167-2188.

Barrett, T.W. (1976) BioSystems 8, 103-109.

Bitter, T. and Muir, H.M. (1962) Anal. Biochem. 4, 330-334.

Blout, E. and Stryer, L. (1959) Proc. Natl. Acad. Sci. USA 45, 1591-1593.

Bree, A. and Lyons, L.E. (1956) J. Chem. Soc. (London) 2658-2670.

Buffington, L.A., Pysh, E.S., Chakrabarti, B. and Balazs, E.A. (1977) J. Am. Chem. Soc. 99, 1730-1734.

Bustamante, C., Tinoco, I., Jr., and Maestre, M.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3568-3572.

Cael, J.J., Winter, W.T. and Arnott, S. (1978) J. Mol. Biol. 125, 21-42.

Cantor, C.R. and Schimmel, P.R. (1980) Biophysical Chemistry, W.H. Freeman and Co., San Francisco.

Causley, G.C. and Johnson, W.C. Jr. (1982) Biopolymers 21, 1763-1780.

Chakrabarti, B. (1977) Arch. Biochem. Biophys. 180, 146-150.

Chakrabarti, B. and Balazs, E.A. (1973) J. Mol. Biol. 78, 135-141.

Christner, J.E., Brown, M.L., and Dziewiatkowski, D.D. (1979a) J. Biol. Chem. 254, 4624-4630.

Christner, J.E., Brown, M.L., and Dziewiatkowski, D.D. (1979b) J. Biol. Chem. 254, 12303-12305.

Cleland, R.L. (1979b) Biochem. Biophys. Res. Commun. 87, 1140-1145.

Cleland, R.L. (1979a) Biopolymers 18, 2673-2681.

Cleland, R.L. and Wang, J.L. (1970) Biopolymers 9, 799-810.

Cowman, M.K., Balazs, E.A., Bergmann, C.W. and Meyer, K. (1981) Biochemistry 20, 1379-1385.

Cowman, M.K., Bush, C.A. and Balazs, E.A. (1983) Biopolymers 22, 1319-1334.

Cowman, M.K., Cozart, D., Nakanishi, K. and Balazs, E.A. (1984) Arch. Biochem. Biophys. 230, 203-212.

Darke, A., Finer, E.G., Moorhouse, R. and Rees, D.A. (1975) J. Mol. Biol. 99, 477-486.

Dea, I.C.M., Moorhouse, R., Rees, D.A., Arnott, S., Guss, J.M. and Balazs, E.A. (1973) Science 179, 560-562.

Deming, S.N. and Morgan, S.L. (1973) Anal. Chem. 45, 278A-283A.

Derby, M.A. (1978) Dev. Biol. 66, 321-336.

Dische, Z. (1947) J. Biol. Chem. 167, 189-198.

Faltz, L.L., Caputo, C.B., Kimura, J.H., Schrode, J. and Hascall, V.C. (1979) J. Biol. Chem. 254, 1381-1387.

Fisher, M. and Solursh, M. (1977) J. Embryol. Exp. Morphol. 42, 195-207.

Forsythe, G.E., Malcolm, M.A. and Moler, G.B. (1977) in Computer Methods for Mathematical Computations, Prentice-Hall, Inc., New Jersey.

Fransson, L. (1974) Carbohydr. Res. 36, 339-348.

Gaskin, F. and Yang, J.T. (1971) Biopolymers, 10, 631-645.

Gill, S.J. and Thompson, D.S. (1967) Proc. Natl. Acad. Sci. USA 57, 562-566.

Golischowski, A.M., King, S.R. and Mascaro, K. (1980) Biochem. J. 192, 1-8.

Goldberg, R.L. and Toole, B.P. (1984) J. Cell Biol. 99, 2114-2122.

Guggenheim, E.A. (1955) Boltzmann's Distribution Law, Interscience, N.Y.

Guss, J.M., Hukins, D.W.L., Smith, P.J.C., Winter, W.T., Arnott, S., Moorhouse, R. and Rees, D.A. (1975) J. Mol. Biol. 95, 359-384.

Hardingham, T.E. and Muir, H. (1972) Biochim. Biophys. Acta, 279, 401-405.

Hascall, V.C. and Heinegard, D. (1974a) J. Biol. Chem. 249, 4242-4249.

Hascall, V.C. and Heinegard, D. (1974b) J. Biol. Chem. 249, 4232-4241.

Hascall, V.C. (1977) J. Supramol. Struc. 7, 101-120.

Hayashi, Y., Teramoto, A., Kawahara, K. and Fujita, H. (1969) Biopolymers, 8, 403-420.

Heatley, F., Scott, J.E., Jeanloz, R.W. and Walker-Nasir, E. (1982) Carbohydr. Res. 99, 1-11.

Hennessey, J.P., Jr. and Johnson, W.C., Jr. (1981) Biochemistry 20, 1085-1094.

Hook, M., Lindahl, U., Hallen, A. and Backstrom, G. (1975) J. Biol. Chem. 250, 6065-6071.

Johnson, W.C., Jr. (1971) Rev. of Sci. Instru. 42, 1283-1286.

Kittel, C. (1969) Thermal Physics, John Wiley and Sons, Inc., N.Y.

Lantzke, I.R. (1973) in Physical Chemistry of Organic Solvent Systems, Covington, A.K. and Dickinson, T. eds., Plenum Press, N.Y., Chapter 4, part 1.

Laurent, T.C. (1966) Fed. Proc. 25, 1037-1038.

Laurent, T.C. (1970) In Chemistry and Molecular Biology of the Intercellular Matrix (Balazs, E.A., ed.) vol. 2, 703-732, Academic Press, N.Y.

Lindahl, U., Hook, M., Backstrom, G., Jacobsson, I., Riesenfeld, J., Malmstrom, A., Roden, L and Feingold, D.S. (1977) Fed. Proc., Fed. Amer. Soc. Exp. Biol. 36, 19-24.

Lindahl, U. and Hook, M. (1978) Ann. Rev. Biochem. 47, 385-417.

Lindahl, U., Backstrom, G., Hook, M., Thunberg, L., Fransson, L. and Linker, A. (1979) Proc. Natl. Acad. Sci. USA 76, 3198-3202.

Listowsky, I., Englard, S. and Avigad, G. (1969) Biochemistry 8, 1781-1785.

McMurtrey, J., Radhakrishnamurthy, B., Dalferes, E.R., Jr., Berenson, G.S., and Gregory, J.D. (1979) J. Biol. Chem. 254, 1621-1626.

Martin, F.H., Uhlenbeck, O.C. and Doty, P. (1971) J. Mol. Biol. 57, 201-215.

Mason, R.M., d'Arville, C., Kimura, J.H. and Hascall, V.C. (1982) Biochem. J. 207, 445-457.

Mathews, M.B. (1975) Molecular Biology, Biochemistry, and Biophysics vol. 19: Connective Tissue. Macromolecular Structure and Evolution, Springer, N.Y.

Meyer, K. (1958) Fed. Proc., Fed. Amer. Soc. Exp. Biol. 17, 1075-1077.

Mitra, A.K., Arnott, S., Atkins, E.D.T. and Isaac, D.H. (1983a) J. Mol. Biol. 169, 873-901.

Mitra, A.K., Arnott, S. and Sheehan, J.K. (1983b) J. Mol. Biol. 169, 813-827.

Mitra, A.K., Raghunathan, S., Sheehan, J.K., and Arnott, S. (1983c) J. Mol. Biol. 169, 829-859.

Morden, K.M., Chu, Y.G., Martin, F.H. and Tinoco, I., Jr. (1983) Biochemistry 22, 5557-5563.

Morgan, S.L. and Deming, S.N. (1974) Anal. Chem. 46, 1170-1181.

Morris, E.R., Rees, D.A. and Robinson, G. (1980c) J. Mol. Biol. 138, 349-362.

Morris, E.R., Rees, D.A., Robinson, G. and Young, G.A. (1980a) J. Mol. Biol. 138, 363-374.

Morris, E.R., Rees, D.A. and Welsh, E.J. (1980b) J. Mol. Biol. 138, 383-400.

Nakamoto, K., Suga, H., Seki, S., Teramoto, A., Norisuye, T. and Fujita, H. (1974) Macromolecules, 7, 784-789.

Napier, M.A. and Hadler, N.M. (1978) Proc. Natl. Acad. Sci. USA, 75, 2261-2265.

Nelder, J.A. and Mead, R. (1965) Comput. J. 7, 308-313.

Nelson, J.W., Martin, F.H. and Tinoco, I., Jr. (1981) Biopolymers, 20, 2509-2531.

Neville, D.M., Jr. and Bradley, D.F. (1961) Biochim. Biophys. Acta 50, 397-399.

Nordenman, B., Danielsson, A. and Bjork, I. (1978) Eur. J. Biochem. 90, 1-6.

Park, J.W. and Chakrabarti, B. (1977a) Biopolymers 16, 2807-2809.

Park, J.W. and Chakrabarti, B. (1978a) Biopolymers 17, 1323-1333.

Park, J.W. and Chakrabarti, B. (1978b) Biochim. Biophys. Acta 544, 667-675.

Park, J.W. and Chakrabarti, B. (1978c) Biochim. Biophys. Acta 541, 263-269.

Potenzone, R. and Hopfinger, A.J. (1975) Carbohydr. Res. 40, 323-336.

Preston, B.N., Davies, M. and Ogston, A.G. (1965) Biochem. J. 96, 449-474.

Rees, D.A. and Welsh, E.J. (1977) Angew. Chem. Int. Ed. Engl. 16, 214-224.

Rees, D.A. (1969) Advan. Carbohydr. Chem. Biochem. 24, 267-332.

Robinson, G., Morris, E.R. and Rees, D.A. (1980) J. Chem. Soc., Chem. Commun. 152-153.

Salzer, H.E., Richards, C.H. and Arsham, I. (1958) Table for the Solution of Cubic Equations, McGraw Hill Book Co., Inc., New York.

Scott, J.E. (1968) Biochim. Biophys. Acta 170, 471-473.

Scott, J.E. and Heatley, F. (1982) Biochem. J. 207, 139-144.

Scott, J.E., Heatley, F. and Hull, W.E. (1984) Biochem. J. 220, 197-205.

Scott, J.E. and Heatley, F. (1979) Biochem. J. 181, 445-449.

Scott, J.E., Heatley, F., Moorcroft, D. and Olavesen, A.H. (1981) Biochem. J. 199, 829-832.

Scott, J.E. and Tigwell, M.J. (1978) Biochem. J. 173, 103-114.

Shaw, B.R. and Schurr, J.M. (1975) Biopolymers, 14, 1951-1985.

Sheehan, J.K., Gardner, K.H. and Atkins, E.D.T. (1977) J. Mol. Biol. 117, 113-135.

Sheehan, J.K., Atkins, E.D.T. and Nieduszynski, I.A. (1975) J. Mol. Biol. 91, 153-163.

Stone, A.L. (1969) Biopolymers 7, 173-188.

Stone, A.L. (1964) Biopolymers 2, 315-325.

Stone, A.L. (1965) Biopolymers 3, 617-624.

Stone, A.L. (1971) Biopolymers 10, 739-751.

Sugahara, K., Schwartz, N.B. and Dorfman, A. (1979) J. Biol. Chem. 254, 6252-6261.

Swann, D.A. and Caulfield, J.B. (1975) Connective Tissue Res. 4, 31-39.

Toole, B.P., Biswas, C. and Gross, J. (1979) Proc. Natl. Acad. Sci. USA 76, 6299-6303.

Turley, E.A. and Roth, S. (1979) Cell, 109-115.

Turley, E.A. (1982) Biochem. Biophys. Res. Comm. 108, 1016-1024.

Turley, E.A. and Roth, S. (1980) Nature, 283, 268-271.

Turley, E.A., Bowman, P. and Kytryk, M.A. (1985) J. Cell Sci. 78, 133-145.

Turner, R.E. and Cowman, M.K. (1985) Arch. Biochem. Biophys. 237, 253-260.

Underhill, C.B. (1982) J. Cell Sci. 56, 177-189.

Wasteson, A., Westermark, B., Lindhal, U. and Ponten, J. (1973) Int. J. Cancer 12, 169-178.

Welsh, E.J., Rees, D.A., Morris, E.R. and Madden, J.K. (1980) J. Mol. Biol. 138, 375-382.

Williams, D.E. and Reisfeld, R.A. (1964) Ann. N.Y. Acad. Sci. 121, 373-381.

Winter, W.T., and Arnott, S. (1977) J. Mol. Biol. 117, 761-784. 169, 829-859.

Winter, W.T., Smith, P.J.C. and Arnott, S. (1975) J. Mol. Biol. 99, 219-235.

Zaks, Rodney (1979a) Programming the 6502, Sybex Inc., Berkeley, CA.

Zaks, Rodney (1979b) 6502 Applications Book, Sybex Inc., Berkeley, CA.

Zimm, B.H., Doty, P. and Iso, K. (1959) Proc. Natl. Acad. Sci. USA 45, 1601-1607.

Zimm, B.H. and Bragg, J.K. (1959) J. Chem. Phys. 31, 526-535.

Zimm, B.H. and Crothers, D.M. (1962) Proc. Natl. Acad. Sci. USA 48, 905-911.

APPENDICES

APPENDIX I: BOLTZMANN TERMS

The following discussion can be found in Chapter 6 of Kittel (1969). It is presented here to clarify the relationship between a Boltzmann term and the statistical probability of the corresponding system state.

Consider a small system in thermal and diffusive molecular contact with a much larger reservoir, the system and reservoir comprising a large closed system with energy U_o and number of molecules N_o. If the small system is in a defined state with N_1 molecules and energy E_1, the probability $P(N_1,E_1)$ that it is in that state will be proportional to the number of accessible states of the reservoir, which has $N_o - N_1$ molecules and energy $U_o - E_1$.

$$P(N_1,E_1) \propto g(N_o-N_1,U_o-E_1) \qquad \text{(AI.1)}$$

Here $g(N,U)$ is the number of states accessible to a closed system with energy U and number of particles N. The relative probability for occupying two different states of the small system is equal to the ratio of the corresponding numbers of accessible states of the reservoir.

$$\frac{P(N_1,E_1)}{P(N_2,E_2)} = \frac{g(N_o-N_1,U_o-E_1)}{g(N_o-N_2,U_o-E_2)} \qquad \text{(AI.2)}$$

The g's are very large numbers. It is easier to deal with the entropy, $\sigma(N,U)$, defined as $\log(g)$.

$$g(N_o,U_o) = e^{\sigma(N_o,U_o)} \tag{AI.3}$$

The relative probability of two states of the small system may be written

$$\frac{P(N_1,E_1)}{P(N_2,E_2)} = \frac{e^{\sigma(N_o-N_1,U_o-E_1)}}{e^{\sigma(N_o-N_2,U_o-E_2)}} = e^{\Delta\sigma} \tag{AI.4}$$

where $\Delta\sigma = \sigma(N_o-N_1,U_o-E_1) - \sigma(N_o-N_2,U_o-E_2)$. The reservoir is assumed very large relative to the small system, so that σ may accurately be expressed as the first order terms in a power series.

Using $$f(x+a) = f(x) + a\frac{df}{dx} + \frac{a^2}{2!}\frac{d^2f}{dx^2} + \ldots. \tag{AI.5}$$

σ for the reservoir becomes

$$\sigma(N_o-N,U_o-E) = \sigma(N_o,U_o) - N\left(\frac{\partial\sigma}{\partial N_o}\right)_{U_o} - E\left(\frac{\partial\sigma}{\partial U_o}\right)_{N_o} + \text{higher order terms}$$

$$\text{and } \Delta\sigma = -(N_1-N_2)\left(\frac{\partial\sigma}{\partial N_o}\right)_{U_o} - (E_1-E_2)\left(\frac{\partial\sigma}{\partial U_o}\right)_{N_o} \tag{AI.6}$$

We define temperature τ and chemical potential μ

$$\frac{1}{\tau} \equiv \frac{\partial\sigma}{\partial U}_{N}, \quad -\frac{\mu}{\tau} \equiv \frac{\partial\sigma}{\partial N}_{U} \tag{AI.7}$$

$$\text{so that } \Delta\sigma = \frac{(N_1-N_2)\mu}{\tau} - \frac{(E_1-E_2)}{\tau} \tag{AI.8}$$

Note that $\tau = kT$ where k is Boltzmann's constant, T is absolute temperature.

Now $$\frac{P(N_1,E_1)}{P(N_2,E_2)} = \frac{e^{(N_1\mu - E_1)/\tau}}{e^{(N_2\mu - E_2)/\tau}} \qquad \text{(AI.9)}$$

Here $e^{(N\mu - E)/\tau}$ is called a Gibbs factor. If the number of particles in the system is fixed, $N_1 = N_2$ and the ratio reduces to

$$\frac{P(E_1)}{P(E_2)} = \frac{e^{-E_1/\tau}}{e^{-E_2/\tau}} \qquad \text{(AI.10)}$$

where the term $e^{-E/\tau}$ is called a Boltzmann factor.

APPENDIX II: CUBIC ROOT ALGORITHM

The following algorithm for solving a polynomial equation of degree 3 is taken from Salzer et al. (1958). Given an equation $ax^3 + bx^2\ cx + d = 0$ to be solved for x, first eliminate the term in x^2 with the substitution $x = y - b/3a$, to obtain a new equation $a'y^3 + b'y + c' = 0$, with

$$a' = a\ ,\ b' = c - \frac{b^2}{3a},\ c' = d - \frac{bc}{3a} + \frac{2b^3}{27a^2}$$

The roots of this cubic are $\frac{-c}{b} f_i(\theta)$ $i = 1, 2, 3$

where $\theta = \frac{a'c'^2}{b'}$ and

$$f_1(\theta) = \frac{\left\{1 + \sqrt{1 + \frac{4}{27\theta}}\right\}^{1/3} + \left\{1 - \sqrt{1 - \frac{4}{27\theta}}\right\}^{1/3}}{(2\theta)^{1/3}}$$

$$f_2(\theta) = \frac{-f_1(\theta)}{2} + \sqrt{\frac{f_1(\theta)^2}{4} - \frac{1}{\theta f_1(\theta)}}$$

$$f_3(\theta) = \frac{-f_1(\theta)}{2} - \sqrt{\frac{f_1(\theta)^2}{4} - \frac{1}{\theta f_1(\theta)}}$$

For $\frac{-4}{27} < \theta < 0$, all roots are real.

For $\theta > 0$ or $\theta < \frac{-4}{27}$, there is one real and two complex roots.

When $\theta = a' = 0$, the only root is $\frac{-c'}{b'}$.

When $\theta = c' = 0$, the three roots are $0, \pm \sqrt{\frac{-b'}{a'}}$.

For $\theta = \frac{-4}{27}$ there are two roots, $\frac{3c'}{b'}, \frac{-3c'}{2b'}$.

APPENDIX III: DIRECTED SEARCH ALGORITHM

The following program contains the logic used to direct the simplex as it seeks the minimum error solution in the vector space of model parameters. It is written in the C programming language (Kernighan and Ritchie, 1972), utilizing a function library written by Rand Dow (Rand Enterprises) for Digital Equipment Corp. software compatability.

The rules for simplex movement are outlined in Morgan and Deming (1974). Briefly, the simplex of n+1 vertices in n-dimensional space has its worst vertex discarded. The reflection of this vertex through the centroid (average) of remaining simplex elements is investigated, and accepted as the new vertex if its error is not an extreme. The vector from the centroid to the vertex can be expanded or contracted depending upon whether the response at the vertex is extremely favorable (expansion) or unfavorable (contraction, inverted contraction). Boundaries on the space to be searched are established by selecting limits for the n dimensions and comparing the new vertex coordinates with them. The vertex is assigned an arbitrary, high error when the limits are exceeded. The simplex movement continues until it contracts to the point where many consecutive new vertices are within a previously established small distance (tolerance) of one another. It is conceivable that the simplex will begin to circle in the vector space without converging on a solution, and new vertices are compared with up to 10 previously discarded vertices in order to detect this condition.

The two subroutines in the program which will require editing for

application to different problems are calpt, simin and sinit. The routine calpt contains the algorithm for calculating values to be compared to experimental observables. Routine simin is for data input, and sinit is for simplex initialization. Initialization is quite important here, as we use it to further limit the vector space in which we search for solutions. This does not mean that we are choosing the solution before we begin--the most important element in the initialization process is still random number generation. Because of limitations in the calculating ability of the computer, we must sacrifice some parameter independence in order to search a generous domain without placing the original simplex such that its vertices are all essentially equally in error.

```
/*****************************************************************
simlog.c contains the logic for the operation of simplexing routines under
RT-11.  the only parameter which may require changing for a particular problem
is the value of NV (number of simplex variables).  NV is not simply passed as
a variable from the input subroutine because it is used to dimension various
arrays, and assigning space for the arrays during program execution (malloc)
could become a bit messy.  CEBUG, EBUG, and DEBUG can all be defined if
debugging output is needed.  TALK, if defined, will sound the terminal bell
every time a simplex answer is found.  Note that if the program is terminated
with ctrl/C, an intermediate execution file named "simplex.mon" will be created
on dk:.  When restarted, the program will look for this file in an attempt to
resume execution where it left off, and so the file must be renamed or deleted
to restart the program with new parameters.
                                        P.S.    7/5/84
*****************************************************************/


#include <rt11.h>
#include <stdio.h>
#include <math.h>
/**include <tsx.h>*/

/**define CEBUG 1               /* primary debugging statements */
/**define EBUG 1                /* for rank debugging */
#define DEBUG 1         /* additional debugging statements */
/**define TALK 1        /* sounds terminal bell when answer found */
/**define INITBUG 1*/
/**define TRANSBUG 1*/

#define CTRLC 1         /* to intercept ctrl/C termination */

#define PASS1 3240000   /* 15 hr in 60 Hz */
#define PASS2 1944000   /* 24-15 hr in 60 Hz */

#define NV 7            /* # of simplex variables */

double vdel[NV] = {0,0};
double cent[NV] = {0,0};
double test[NV] = {0,0};

int rank[NV+12] = {0,0};
int nruns = 0;
int match = 0;

struct simp
```

```
        {
        double var[NV];
        double sum;
        };


struct simp simplx[NV+12] = {0,0};

int args = FALSE;               /* no system arguments */
int circnt =0;
int monitor = 0;
int next = 0;
int dbgflg = 0;

double tolerance = 0;
double toler1 = 0;
double toler2 = 0;

FILE *simpin = 0;
FILE *output = 0;

FILE *dbgout = 0;

char simpar[20] = "dk:simplx.par";
char dbgname[20] = "tt:";       /* debugging output device */
main()

{

double tsum,ttsum,factor;
double elapse,sum();

int start,stop;
int i,j,k,rankmp,num;
int heyyou;

int tmprnk[NV+2];
double tmpsum[NV+2];

long tstart,tcurrent;


        simin();        /* get run parameters, input data */

/*      if(!Tsxpriv)
```

```
                printf(" I can't lock into memory from this terminal!\n");
        else Tsxrtlock;
*/
        monitor = 0;    /* tracks # of cycles in search */

#ifdef CTRLC
        heyyou = 0;     /* to store CTRL/C's */

        Scca(&heyyou);  /* suppress immediate CTRL/C abort */
#endif
#ifdef DEBUG
        dbgout = fopen(dbgname,"r+");
#endif
        num = NV;

/*************************************************************************
simplx.mon is a file created when the routine is terminated with CTRL/C.
it contains all of the information needed to restart the program where it left
off.  thus, the monitor file facilitates interrupting program execution with-
out losing valuable run time toward a 'hard to get' answer.  the monitor file
must be deleted or renamed to restart an interrupted program with new run
parameters.
*************************************************************/

        if(simpin = fopen("dk:simplx.mon","r"))
                {printf("restarting from monitor file\n\n");
                fscanf(simpin,"%s%d%d%d",simpar,&mruns,&monitor,&next);
                for(j=0;j<NV+12;j++)
                        {for(i=0;i<NV;i++)
                                fscanf(simpin,"%f",&simplx[j].var[i]);
                        fscanf(simpin,"%f",&simplx[j].sum);
                        }
#ifdef DEBUG
                fprintf(dbgout,"ranks\n");
#endif
                for(i=0;i<NV+12;i++)
                        {fscanf(simpin,"%d",&rank[i]);
#ifdef DEBUG
                        fprintf(dbgout,"%d\n",rank[i]);
#endif
                        }
                fscanf(simpin,"%f%d",&tolerance,&circnt);
                fclose(simpin);
                Gtim(&tstart);  /* record time of answer cycle start */
                goto bypass;    /* bypass usual initialization */
```

```
            }

/*************************************************************
        THIS LEVEL 2 LOOP CONTINUES UNTIL THE NUMBER OF ANSWERS DESIGNATED
        IN THE PARAMETER OR MONITOR FILE IS FOUND
*************************************************************/
while(nruns>0)
{
        /***********************************************
                THIS IS THE USUAL INITIALIZATION SECTION
        ***********************************************/
start:  circnt = 0;

        tolerance = toler1;     /* 1st termination tolerance */

        next = 0;
        monitor = 0;

        Gtim(&tstart);          /* time of answer cycle start */

        for(j=0;j<=NV;j++)      /* initialize simplex */
                sinit(&j);
        for(i=0;i<NV+12;i++)
                {rank[i] = i;
                simplx[i].sum = 1000000.0;
                }
        /*****************************************************
                THIS LEVEL 3 LOOP CONTINUES UNTIL THE SIMPLEX MATCHES SOME
                PREVIOUS POINT IN THE SERIES 8 CONSECUTIVE TIMES, EITHER BY
                MOVING THROUGH THE VECTOR SPACE IN A LOOP (CIRCLE SIZE <=12)
                OR BY BECOMING SMALL ENOUGH (WITHIN ANSWER TOLERANCE)
        *****************************************************/

bypass: while(circnt<8)
        {

#ifdef DEBUG
        for(i=0;i<=NV;i++)
                {fprintf(dbgout,"\nsimplex point %d ",i);
                for(j=0;j<NV;j++)
                        fprintf(dbgout,"%s%f",j%5? "   " : "\n\t\t",
                                simplx[i].var[j]);
                fprintf(dbgout,"\n");
                }
#endif
```

```
        if(!next)       /* next flag is 0 only to begin a search, if the
                           simplex has translated, or if it has been reset
                           in the monitor file */
                {

                /* initialize rank shuffle area */
                for(i=0;i(=NV;i++)
                        {tmprnK[i] = NV+1;
                        tmpsum[i] = 1000000.0;
                        }
#ifdef DEBUG
                for(i=0;i(=NV+11;i++)
                        {fprintf(dbgout,"\nsimplex point %d ",i);
                        for(j=0;j(NV;j++)
                                fprintf(dbgout,"%s%f",j%5? "  " : "\n\t\t",
                                        simplx[i].var[j]);
                        fprintf(dbgout,"\n");
                        }
#endif
#ifndef TRANSBUG
                dbgflg = 0;
#endif
#ifdef INITBUG
                dbgflg = 1;
#endif
#ifdef DEBUG
                fprintf(dbgout,"\n\n****REORDER SIMPLEX POINTS*********\n");
#endif

                /* here we make sure that all simplex points have an error of
                   ( 1000000., or they are reinitialized */
                for(i=0;i(=NV;i++)
                        {
                        while((tsum = sum(simplx[rank[i]].var,dbgflg)) == 1000000.0)
                                sinit(&rank[i]);

                        /* now the simplex point is ranked according to its
                           error (sum()) */
                        j = NV;
                        while((tsum(tmpsum[j])&&(j)=0))
                                {
#ifdef EBUG
                                fprintf(dbgout,"%f ( %f so rank %d (simplex pt.",
                                        tsum,tmpsum[j],j);
                                fprintf(dbgout,"%d) pushed to %d\n",tmprnK[j],j+1);
```

```
#endif
                                tmprnk[j+1] = tmprnk[j];
                                tmpsum[j+1] = tmpsum[j--];
                                }
                        tmprnk[++j] = rank[i];
                        tmpsum[j] = tsum;
#ifdef EBUG
                        fprintf(dbgout,"simplex point %d in position %d",
                                rank[i],j);
                        fprintf(dbgout," with sum of %f\n\n",tmpsum[j]);
#endif
                        }
                for(i=0;i(=NV;i++)
                        {rank[i] = tmprnk[i];
                        simplx[rank[i]].sum = tmpsum[i];
                        }
                next = NV;      /* next traces next simplex point to be */
                }               /* replaced (its rank)                  */

#ifdef INITBUG
        dbgflg = 0;
#endif
#ifdef CEBUG
        dbgflg = 1;
        fprintf(dbgout,"summary of current rank of simplex points\n");
        for(i=0;i(=NV;i++)
                {fprintf(dbgout,"#%d with sum %f",
                        rank[i],simplx[rank[i]].sum);
                for(j=0;j(NV;j++)
                        fprintf(dbgout,"%s%f",j%5? "  " : "\n\t\t",
                                simplx[rank[i]].var[j]);
                fprintf(dbgout,"\n");
                }
#endif
        /*********************************************
                GET NEW SIMPLEX POINT
        *********************************************/
        /* here we calculate the centroid of the simplex minus its worst point
        (highest rank value) */
        zero(cent,sizeof(cent[0])*NV);  /* initialize */
        for(i=0;i(=NV;i++)
                arrdif(cent,simplx[rank[i]].var,cent,&num,"+");
        arrdif(cent,simplx[rank[next]].var,cent,&num,"-");
        factor = 1.0/(double)NV;
```

```
        multar(cent,&factor,&num);

        /* now we move that worst data point (reflection through centroid) */
        arrdif(simplx[rank[next]],var,cent,vdel,&num,"-");
        arrdif(cent,vdel,test,&num,"-");

#ifdef CEBUG
        fprintf(dbgout,"new test point follows\n");
        for(k=0;k<NV;k++)
                fprintf(dbgout,"value %d is %f\n",k,test[k]);
        fprintf(dbgout,"debug flag is %d",dbgflg);
#endif

        tsum = sum(test,dbgflg);        /* get new error */

        for(i=0;i<=NV;i++)      /* and rank new point */
                if(tsum<simplx[rank[i]],sum)break;
#ifdef CEBUG
        fprintf(dbgout,"new point should land in position %d with sum %f\n",
                i,tsum);
#endif
        /*********************************************************
                HERE WE BRANCH ON LEVEL 4 DEPENDING UPON THE ERROR
                SUM OF THE NEW SIMPLEX POINT
        *********************************************************/
/*switch(i)     {
        case 0:*/

        if(i==0)        /* the new point is now the best one, so we want to */
                {       /* look further in its direction                    */
#ifdef CEBUG
                fprintf(dbgout,"new point is best one, expand its vdel\n");
#endif
                if(next < NV)next++;
                factor=0.5;
                multar(vdel,&factor,&num);      /* expand delta vector */
                arrdif(test,vdel,test,&num,"-");
                if((ttsum = sum(test,dbgflg)) >= tsum)
                        /* check error, and if greater, go back    */
                        {
#ifdef CEBUG
                        fprintf(dbgout,"expansion fails, contract back\n");
#endif
                        arrdif(test,vdel,test,&num,"+");
                        }
```

```
                else tsum = ttsum;

/*              break;*/
                }
        /*******************************************************/
/*      case NV:*/

        else if(i==next)        /* new point still worst in simplex, so */
                {               /* we look away from it instead         */
#ifdef CEBUG
                fprintf(dbgout,"new point still worst in set, contract its vdel\n");
#endif
                if(next < NV)next++;
                factor=0.5;
                multar(vdel,&factor,&num);      /* reverse delta vector */
                arrdif(test,vdel,test,&num,"+");
                if((ttsum = sum(test,dbgflg)) >= tsum)
                        /* check error, and if greater, reverse vector */
                        {
#ifdef CEBUG
                        fprintf(dbgout,"contraction fails, expand back\n");
#endif
                        arrdif(test,vdel,test,&num,"-");
                        }
                else    {tsum = ttsum;
                        for(i=0;i<=NV;i++)
                                if(tsum<simplx[rank[i]].sum)break;
                        }

/*              break;*/

                }
        /*******************************************************/
/*      case NV+1:*/

        else if(i > next)       /* moving simplex point makes it terrible, */
                {               /* so look in other direction           */
#ifdef CEBUG
                fprintf(dbgout,"new point REALLY bad, invert and contract its vdel\n");
#endif
                factor=1.5;
                multar(vdel,&factor,&num);
                arrdif(test,vdel,test,&num,"+");
                if((ttsum = sum(test,dbgflg)) >= simplx[rank[next]].sum)
                        {
```

```
#ifdef CEBUG
                        fprintf(dbgout,"contraction fails, expand instead\n");
#endif
                        factor = 2.0/3.0;
                        multar(vdel,&factor,&num);
                        arrdif(test,vdel,test,&num,"+");
                        if((ttsum = sum(test,dbgflg)) >= simplx[rank[next]].sum)
                                {       /* can't improve point on this line */
#ifdef CEBUG
                                fprintf(dbgout,"must resort to RULE 2\n");
#endif
                                next--; /* so replace 2nd worst point */
                                }
                        else    {tsum = ttsum;
                                if(next < NV)next++;
                                for(i=0;i<=NV;i++)
                                        if(tsum<simplx[rank[i]].sum)break;
                                }
                        }
                else    {tsum = ttsum;
                        if(next < NV)next++;
                        for(i=0;i<=NV;i++)
                                if(tsum<simplx[rank[i]].sum)break;
                        }

/*              break;*/

                }

        else if(next < NV)next++;

#ifdef DEBUG
        fprintf(dbgout,"\t\t\t\t\t\t\tnext is now %d\n",next);
#endif

        /*****************************************************
                THROW AWAY OLD POINT
        *****************************************************/
        if(next==NV)
                {
#ifdef CEBUG
                fprintf(dbgout,"new simplex point will slip into position %d\n\n",i);
#endif
                ranktmp = rank[NV+1];  /* insert new point in ranking */
                for(j=NV+1;j>i;j--)
```

```
                        {rank[j] = rank[j-1];
#ifdef EBUG
                        fprintf(dbgout,"simplex pt %d at position %d moved to %d\n",rank[j-1],j-1,j);
#endif
                        }
                rank[j] = ranktmp;
#ifdef EBUG
                fprintf(dbgout,"simplex pt %d at rank %d with values of\n",
                        ranktmp,j);
#endif
                for(j=0;j<NV;j++)
                        {simplx[ranktmp].var[j]=test[j];
#ifdef EBUG
                        fprintf(dbgout," %f ",simplx[ranktmp].var[j]);
#endif
                        }
                simplx[ranktmp].sum=tsum;
#ifdef EBUG
                fprintf(dbgout,"\nand sum of %f\n\n",simplx[ranktmp].sum);
#endif
                compare();

                /*****************************************
                        START SECOND PHASE OF (TUKEY) SEARCH IF DESIRED
                *****************************************/
                if(circnt == 8 && (tolerance-toler2) )
                        {tolerance = toler2;
#ifdef DEBUG
                        fprintf(dbgout,"tolerance now %f\n",tolerance);
#endif
                        circnt = 0;
                        next = 0;
                        }

                }

        /*************************************************
                LEVEL 4 LOOP TRANSLATES STALLED SIMPLEX TO PLACE BEST POINT
                (WHICH IS ALSO RETAINED) AT THE 'PSEUDO'-CENTER
        *************************************************/
        if(next==(NV-2))
                {
#ifdef  CEBUG
                fprintf(dbgout,"\n\n**********TRANSLATING*SIMPLEX******\n");
                fprintf(dbgout,"moving simplex over simplx[%d]\n",rank[0]);
```

```
#endif
                for(j=0;j(=NV;j++)
                        arrdif(cent,simplx[rank[j]].var,cent,&num,"+");
                factor = 1/(double)(NV+1);
                multar(cent,&factor,&num);
                arrdif(simplx[rank[0]].var,cent,vdel,&num,"-");
                for(j=1;j(NV;j++)
                        arrdif(simplx[rank[j]].var,vdel,simplx[rank[j]].var,&num,"+");
                next = 0;
                }
        /* check search time limitation */
        Gtim(&tcurrent);
#ifdef DEBUG
        fprintf(dbgout,"\ndelta time is %ld\n\n",tcurrent-tstart);
#endif
/*      if(tcurrent)tstart && (tcurrent-tstart))PASS1)break;
        if(tstart)tcurrent && (tstart-tcurrent)(PASS2)break;
                /* break after 5 hours */
#ifdef CTRLC
        /* break out if ctrl c from terminal */
#ifdef DEBUG
        fprintf(dbgout,"\nheyyou is %d\n",heyyou);
#endif
        if(heyyou)
                {simcut();
                return(TRUE);
                }
#endif
        /* occasional monitor of simplex travel */
        if(!((++monitor)%50))
                {
                simcut();
                elapse = (float)(tcurrent-tstart)/216000.;
                if(elapse ( 0.)elapse += 24.0;
                fprintf(dbgout,"elapsed time in hours: %6.3f\n",elapse);
                for(i=0;i(NV;i++)
                        fprintf(dbgout,"%f%s",simplx[rank[0]].var[i],
                                (((i+1)%5)==0)? "\n\t" : "  ");
                fprintf(dbgout," sum = %f\n",simplx[rank[0]].sum);
                }
        }

        /* write answer */
        simdat();
```

```
#ifdef TALK
        printf("%c",'\007');
#endif

#ifdef DEBUG
        fprintf(dbgout,"going back for next run, %d\n",nruns-1);
#endif
        match = 0;
        nruns--;
}

        return(TRUE);

}

/*************************************************************************/

compare()


        /* initialize match parameter */
{
int start,stop,i,j,K;

        start = NV+2;
        stop = NV+11;
        if(match)start = stop = match;
        match = 0;
        /***********************************
                LEVEL 5 LOOP LOOKS FOR MATCHES TO PREVIOUS
                SIMPLEX POINTS
        ***********************************************/
        for(j=start;j(=stop;j++)
                {i = inner(&j);
                if(i==NV)
                        {match = j;
#ifdef DEBUG
                        fprintf(dbgout,"new point matches %d pts previous\n",j-(NV+1));
                        for(K=0;K(NV;K++)
                                {fprintf(dbgout,"%f - %f / %f ( tolerance %f ?\n",simplx[rank[j]],var[K],test[K]
                                test[K],tolerance);
                                }
#endif
                        break;
                        }
```

```
                }
        if(match)circnt++;
        else circnt = 0;
#ifdef DEBUG
        if(match)fprintf(dbgout,"this match has occurred %d times\n",
                circnt);
#endif
        return(TRUE);
}

/*****************************************************/

int inner(j)

int *j;

{

float diff;
int i,k;

        i = 0;
        for(k=0;k<NV;k++)
                {
                diff=abs(simplx[rank[*j]].var[k]-test[k]);
                if(abs(test[k])<1.0 && diff<tolerance)
                        {i++;
#ifdef DEBUG
                        printf("abs(%f) < 1.0 and %f-%f < %f, i = %d\n",
                                test[k],simplx[rank[*j]].var[k],test[k],
                                tolerance,i);
#endif
                        }
                else if(abs(diff/test[k])<tolerance)
                        {i++;
#ifdef DEBUG
                        printf("(%f-%f=%f)/%f < %f, i = %d\n",
                                simplx[rank[*j]].var[k],test[k],diff,
                                test[k],tolerance,i);
#endif
                        }
                }
        return(i);
}
```

```
/* This form of simzip analyzes data at a given temperature and various
oligomer lengths - P.S. */

#include <stdlib.h>
#include <math.h>
#include <stdio.h>

#define NV 5    /* number of simplex variables, must be same as in simlog*/

/**define SIMCUT 1
#define SIMINBUG 1
#define INITBUG 1
#define SUMBUG 1
#define CALBUG 1*/

static int REGISTER = 0;
static int TRIMER = 0;
static int PARALLEL = 0;

static int length = 0;
static double temp = 0.0;
external int rank[],nruns,match,circnt,next,monitor;
external int dbgflg;

external double tolerance,toler1,toler2;

external long simpin,output;

external char simpar[],dbgname[];

struct simp
        {
        double var[NV];
        double sum;
        };
external struct simp simplx[];

static double obs[200] = {0.,};
static double ind[200] = {0.,};
static double res[200] = {0.,};
static double wres[200] = {0.,};

static double zmed = 0.;
static double T[200] = {0.,};
static double s = 0.;
```

```
static double alpha = 0.;
static double Jsder = 0.;
static double Js = 0.;

static double ext[4][NV] = {0.,};

static char outfil[20] = "dk:simzip.out";
static char datfil[20] = "dk:simzip.in";

static int npts = 0;
static int lflag = 0;
static int flag = 0;
static int tukflg = 0;
static int scount = 0;
static int N[200] = {0,};

FILE *data = 0;
external FILE *dbgout;

/*****************************************************************/

double  sum(datarr,dbgflg)

double datarr[NV];
int dbgflg;
{

double sum,rest,ewe;
int j,i,mid,flag;
double calpt();

        sum = 0.0;
#ifdef SUMBUG
        if(dbgflg)
        fprintf(dbgout,"\n\n\n**************SUM*****************\n");
#endif

        if(limit(datarr))return(1000000.0);     /*first check for bounds*/

        scount = 0;
        for(j=0;j<npts;j++)
                {
                res[j] = obs[j] - calpt(ind[j],datarr,N[j],T[j]);
                        /*make residuals*/
#ifdef SUMBUG
```

```
        if(dbgflg)fprintf(dbgout,"residual(%d) is %f\n",j,res[j]);
#endif
                wres[j] = res[j] = abs(res[j]);
                }

        if(!tukflg || (tolerance - toler2))
                {for(i=0;i(npts;i++)            /*npts was scount*/
                        {if()flag == 1)sum += res[i];
                        if()flag == 2)sum += res[i]*res[i];
                        }

#ifdef SUMBUG
        if(dbgflg)
                {fprintf(dbgout,"sum for simplex point is %f\n",sum);
                fprintf(dbgout,"*******ENDSUM**********\n");}
#endif
                if(sum ) 1000000.0)sum = 1000000.0;
                return(sum);
                }

        do
                {flag = 0;
                for(i=1;i(scount;i++)
                        {if(wres[i-1] ) wres[i])
                                {rest = wres[i];
                                wres[i] = wres[i-1];
                                wres[i-1] = rest;
                                flag++;
                                break;}
                        }
                }
        while(flag);

#ifdef SUMBUG
        if(dbgflg)
        for(i=0;i(scount;i++)
                fprintf(dbgout,"residuals %d are %f and %f\n",i,
                        res[i],wres[i]);
#endif

        mid = scount/2;
        zmed = (wres[mid] + wres[scount-1-mid])/2.0;

#ifdef SUMBUG
        printf("Tukey residual median %f = (%f+%f)/2\n",zmed,wres[mid],
```

```
                wres[scount-1-mid]);
#endif

        for(i=0;i(scount;i++)
                {
                if(ewe = res[i]/(6.0*zmed) )= 1.0) wres[i] = 0.0;
                else {wres[i] = res[i]*(1-ewe*ewe);
                        wres[i] *= wres[i]*res[i];}
#ifdef SUMBUG
        printf("residual for point %d is %f\n",i,wres[i]);
#endif
                sum += wres[i];
                }

        sum /= scount;

#ifdef SUMBUG
                printf("Tukey sum for simplex point is %f\n",sum);
                printf("**********ENDSUM**************************\n");
#endif
        if(sum ) 1000000.0)sum = 1000000.0;
        return(sum);
        }

/*-----------------------------------------------------------------*/


double calpt(conc,datarr,N,T)

double datarr[NV],conc,T;
int N;
{

double logZ,frac,gamma;
double CDcal,arg,arg2;
int coeff,coeffc,dummy,mult;
double q,r1,b1,yo,logyo,Jssub;
double x,y;

        arg = (datarr[0]/T-datarr[1])/1.98717;
        if(arg ) 78.0/N)arg = 78.0/N;
        if(arg ( -78.0/N)arg = -78.0/N;
#ifdef CALBUG
        if(dbgflg)fprintf(dbgout,"argument of exponential is %f\n",arg);
#endif
```

```
        s = exp(-arg);
        alpha = 0.4342945*(datarr[2]/T-datarr[3])/1.98717;
#ifdef CALBUG
        if(dbgflg)
                {fprintf(dbgout,"s = exp((%f-%f/%f)/1.98717) = %f\n",datarr[1]
                        ,datarr[0],T,s);
                fprintf(dbgout,"alpha = .4343(-%f+%f/%f)/1.98717 = %f\n\n\n",
                        datarr[3],datarr[2],T,alpha);}
#endif

        arg = conc/N;
/*******logZ = 23.77981+log10(arg);*********************************/
        logZ = log10(arg);

        if((length != N) || (temp != T))
                {
                length = coeffc = N;
                temp = T;
                Jsder = Js = 0;

/*      This calculation of the partition function term coefficients is good
        for chains remaining in register, whether 2 strands in parallel or
        3 strands in parallel or with one antiparallel to the other two  */

        if(REGISTER)
                {if(PARALLEL || TRIMER)
                        for(coeff=1;coeff(=N;coeff++)
                                {
                                Js += coeff;
                                Jsder += coeff*coeffc--;
                                Js *= s;
                                Jsder *= s;
                                }

/*      This calculation of the partition function term coefficients should
        be used for 2 antiparallel strands in register  */

                else    for(coeff=1;coeff(=N;coeff++)
                                {
                                dummy = coeff;
                                if ((dummy/2)*2 ( coeff)
                                        {dummy +=2;
                                        dummy /=2;}
                                Js += dummy;
                                Jsder += dummy*coeffc--;
```

```
                        Js *= s;
                        Jsder *= s;
                        }
                }

        else    {if(TRIMER)
                        {if(!PARALLEL)
/*      This calculation of partition function term coefficients is for
        three strands with one antiparallel to the other two, staggering
        of the ends allowed  */
                        for(coeff=1;coeff(=N;coeff++)
                                {Js += coeff*coeff*coeff;
                                Jsder += coeff*coeff*coeff*coeffc--;
                                Js *= s;
                                Jsder *= s;}


/*      This calculation of partition function term coefficients is for
        three strands  in parallel, staggering of the ends allowed  */
                        else for(coeff=1;coeff(=N;coeff++)
                                {Js += (coeff*coeff*coeff + 2*coeff)/3.;
                                Jsder += (coeff*coeff*coeff + 2*coeff)*coeffc--/3.;
                                Js *= s;
                                Jsder *= s;}
                        }

/*      This calculation of partition function term coefficients is good
        for 2 staggered strands associating in parallel or antiparallel   */

                else for(coeff=1;coeff(=N;coeff++)
                        {Js += coeff*(coeff+1)/2;
                        Jsder += coeff*(coeff+1)*coeffc--/2;
                        Js *= s;
                        Jsder *= s;
                        }
                }
#ifdef CALBUG
        if(dbgflg)fprintf(dbgout,"Js is %f and Jsder %f\n\n",Js,Jsder);
#endif
        }
/*************************************************************
is this a legitimate alternate expression?
        c1 = Z*Z;
        b1 = (2*Z*q+1)/(2*q);
        yo = c1/(b1+sqrt(b1*b1-c1));
```

```
        frac = yo/Z*Jsder/Js/N;
*************************************************************/
        if(!TRIMER)
                {logyo = .60206-alpha+logZ+log10(Js);
#ifdef CALBUG
        if(dbgflg)fprintf(dbgout,"log of yo =%f-%f+%f+%f\n=%f\n",log10(4.0)
                        ,alpha,logZ,log10(Js),logyo);
#endif
                if(logyo > 37.0)logyo = 37.0;
                if(logyo < -37.0)logyo = -37.0;
                yo = exp10(logyo);
                arg = sqrt(1+2*yo);
#ifdef CALBUG
        if(dbgflg)
                {fprintf(dbgout," = %f\n",logyo);
                fprintf(dbgout," sqrt(1+2*%f) = %f\n",yo,arg);
                fprintf(dbgout," 1+%f-%f = ",yo,arg);}
#endif
                frac = 1+yo-arg;
#ifdef CALBUG
        if(dbgflg)fprintf(dbgout," %f\n",frac);
#endif
                arg = frac;
#ifdef CALBUG
        if(dbgflg)fprintf(dbgout,"%f*%f/%f*%f*%d = ",arg,Jsder,yo,Js,N);
#endif
                frac = arg/yo*Jsder/Js/N;
#ifdef CALBUG
        if(dbgflg)fprintf(dbgout,"\nfraction helical (dimer) is %f\n\n",frac);
#endif
                }

        if(TRIMER)
                {conc /= N;
/***************arg = 48.3377 - alpha + 2*log10(conc) + log10(Js);**********/
                arg = .77815 - alpha + 2*log10(conc) + log10(Js);
#ifdef CALBUG
        if(dbgflg)fprintf(dbgout,"48.3377 - %f + 2*%f + %f = %f\n",
                        alpha,log10(conc),log10(Js),arg);
#endif
                arg2 = -arg - 0.52827;
                if(arg2 > 37.0)arg2 = 37.0;
                if(arg2 < -37.0)arg2 = -37.0;
                arg2 = exp10(arg2) + 1;
```

```
                arg2 = sqrt(arg2);
                x = arg2 + 1.;
                y = 1./3.;
                yo = pow(x,y);
#ifdef CALBUG
        if(dbgflg)fprintf(dbgout,"cubert(%e) = %e = yo\n",x,yo);
#endif
                x = 1. - arg2;
                if(x == 0.);
                else if(x > 0.)
                        {
#ifdef CALBUG
        if(dbgflg)fprintf(dbgout,"cubert(%e) = %e\n",x,pow(x,y));
#endif
                        yo += pow(x,y);
                        }
                else if(x < 0.) {x *= -1.;
#ifdef CALBUG
        if(dbgflg)fprintf(dbgout,"minus cubert(%e) = %e\n",x,pow(x,y));
#endif
                                yo -= pow(x,y);
                                }
/****************arg = log10(conc) - log10(Js) - alpha + 23.4788;*************/
#ifdef CALBUG
        if(dbgflg)fprintf(dbgout," - %e = %e\n",(x==0.)? 0.0 : pow(x,y),yo);
#endif
                yo = exp10(arg/3.0) - yo;
                if(yo < 0.)frac = 0.0;

                else    {logyo = log10(yo);
#ifdef CALBUG
        if(dbgflg)fprintf(dbgout,"log of yo is %e\n",logyo);
#endif
                        logyo -= arg/3.0;
                        arg = N;
                        frac = logyo - log10(arg) + log10(Jsder) - log10(Js);
#ifdef CALBUG
        if(dbgflg)fprintf(dbgout,"log of frac is %f - %f + %f - %f\n",logyo,
                                log10(arg),log10(Jsder),log10(Js));
#endif
                        frac = exp10(frac);
#ifdef CALBUG
        if(dbgflg)fprintf(dbgout,"\nfraction helical (trimer) is %f\n\n",
                                frac);
#endif
```

```
                }
            }

/*******CDcal = frac*datarr[6] + (1-frac)*(datarr[4]+datarr[5]*T);**********/
        CDcal = frac*datarr[4] + (1-frac)*(12.02-.04027*T);
        return(CDcal);
}

/*-----------------------------------------------------------------*/

int limit(datarr)

double datarr[NV];

{

int j;

        for(j=0;j<NV;j++)
            {if((datarr[j])<(ext[0][j]))
                {
#ifdef SUMBUG
        if(dbgflg)fprintf(dbgout,"%f < %f so point out of bounds!!\n",
                            datarr[j],ext[0][j]);
#endif
                return(1);
                }
            if((datarr[j])>(ext[1][j]))
                {
#ifdef SUMBUG
        if(dbgflg)fprintf(dbgout,"%f > %f so point out of bounds!!\n",
                            datarr[j],ext[1][j]);
#endif
                return(1);
                }
            }

        return(0);

}

/*-----------------------------------------------------------------*/


simdat()
```

```
{
int i;

        data = fopen(outfil,"amf");

        fprintf(data,"\ncircle size %d: %c cycles\n\tvariables:",
                        match-(NV+1),monitor);
        for(i=0;i(NV-1;i++)
                fprintf(data,"%s%f",(((i+1)%6)==0)? "\n\t\t" : " ",
                        simplx[rank[0]],var[i]);
        fprintf(data,"\n\t\t12.02 -.04027 %f",simplx[rank[0]],var[NV-1]);
        fprintf(data,"\n\t\tL%d_sum %e\n",iflag,simplx[rank[0]],sum);

        if(tukflg)
                {fprintf(data,"Tukey fit:\n");
                for(i=0;i(npts;i++)
                        {fprintf(data,"point %f at %f medians%c",ind[i],
                        res[i]/zmed,(i+1)%2? '\t' : '\n');
                        }
                }

        fclose(data);

}
/*********************************************************************
        ind[] is the independent variable (here it's HA residue [])
        obs[] is the experimentally observed dependent variable (CD)
        npts is the number of data point pairs you wish to use in the fit
        NV is the number of variables needed to fit the data to an
        experimental equation
        datarr is an NV*(NV+1) array of starting values for the fitting
        parameters
        ext[][] is a NV*2 array of the extreme values for the fitting
        parameters .  the minimum is in (1,x) and the maximum
        is in (2,x)
        iflag is the L(1) or L(2) minimization

        the input file (simin.par) is as follows:
        1)      # of simplex runs
        2)    (L)1 or (L)2
        3)   input data file
        4)      output data file
        5)      output device (for debug?)
        6)      simplex variable extrema (low,high)
```

```
        7)      tolerance level for fit (as % of parameters)
        8)      2nd tolerance level (for Tukey switch)
        9)      Tukey flag
        10)     temperature

this subroutine reads in data for the fit to a non-stagger zipper model with
two polymer chains associating, a la Applequist and Damle.
the elements of datarr and ext are:
        0) equilibrium s for random->ordered
        1) beta for chain association equilibrium (beta*s)
        2) random structure CD
        3) ordered structure CD

*************************************************************************/

simin()

{

int i,j;
long simpin;
float dummy;

        if(data = fopen("dK:simplx.mon","r"))
                {fscanf(data,"%s%d",simpar,&nruns);
                fclose(data);}
        else    {printf("enter name of parameter file\n");
                scanf("%s",simpar);}

#ifdef SIMINBUG
        if(dbgflg)
        {fprintf(dbgout,"\n\n\n**********************SIMIN**********************\n");
        fprintf(dbgout,"opening file %s\n\n",simpar);}
#endif
        data=fopen(simpar,"rmf");
        fscanf(data,"%d%d%s%s%s",&nruns,&lflag,datfil,outfil,dbgname);
        for(i=0;i(NV;i++)
                fscanf(data,"%f%f%f%f",&ext[0][i],&ext[1][i],&ext[2][i],
                        &ext[3][i]);
        fscanf(data,"%f%f%d",&toler1,&toler2,&tukflg);
        fclose(data);

#ifdef SIMINBUG
        if(dbgflg)
        {fprintf(dbgout,"%d runs using L%d estimate\n",nruns,lflag);
```

```
        fprintf(dbgout,"extrema for simplex variables follow\n");
        for(i=0;i(NV;i++)
                fprintf(dbgout,"variable %d from %f to %f\n",i+1,ext[0][i],
                        ext[1][i]);
        fprintf(dbgout,"\ntolerance is %f\n\n",toler2);
        fprintf(dbgout,"\nopening %s for data\n\n",datfil);}
#endif

        data = fopen(datfil,"rmf");
        fscanf(data,"%d",&npts);
        for(i=0;i(npts;i++)
                {
                fscanf(data,"%f%f%f%d",&ind[i],&obs[i],&T[i],&N[i]);
#ifdef SIMINBUG
        if(dbgflg)fprintf(dbgout,"%f,%f,%f,%d\n",ind[i],obs[i],T[i],N[i]);
#endif
                }
        fclose(data);

        for(i=0;outfil[i] != '.';i++);
        if(tolower(outfil[i-1]) == 'r')REGISTER = 1;
        else if(tolower(outfil[i-1]) == 's')REGISTER = 0;
        if(tolower(outfil[i-4]) == 'p')PARALLEL = 1;
        else if(tolower(outfil[i-4]) == 'a')PARALLEL = 0;
        if(tolower(outfil[i+1]) == 't')TRIMER = 1;
        else if(tolower(outfil[i+1]) == 'd')TRIMER = 0;

#ifdef SIMINBUG
        if(dbgflg)fprintf(dbgout,"seeding random number generator\n");
#endif
        ransed();

#ifdef SIMINBUG
        if(dbgflg)fprintf(dbgout,"**********ENDSIMIN**********\n");
#endif

}

/*****************************************************************/


int sinit(index)

int *index;
```

```
{
int i;
double wt;
double idiot,random();

#ifdef INITBUG
        if(dbgflg)fprintf(dbgout,"\n\n\n**********SINIT**************\n");
#endif
        for(i=0;i<NV;i++)
                {idiot = ext[0][i] +
                (.01+.98*(wt=random()))*(ext[1][i]-ext[0][i]);
#ifdef INITBUG
        if(dbgflg)fprintf(dbgout,"%f + (.01+.98*%f)(%f - %f) = %f\n",
                        ext[0][i],wt,ext[1][i],ext[0][i],idiot);
#endif
                simplx[*index].var[i] = idiot;

                }
/*      simplx[*index].var[0] = -1+2*(wt=random());
        simplx[*index].var[1] = -21+2*(wt=random());
        if(!TRIMER)
                simplx[*index].var[3] = (simplx[*index].var[2]/295.0-19.0*4.6-
                (wt=random())*9.2);
        if(TRIMER)
                simplx[*index].var[3] = (simplx[*index].var[2]/295.0-239.0+
                (wt=random())*37.0);*/

/*      simplx[*index].var[0] = -870. + 40.*random();
        simplx[*index].var[1] = -2.2 + 0.2*random();
        simplx[*index].var[2] = -20500. + 2000.*random();
        simplx[*index].var[3] = -70. + 10.*random();
        simplx[*index].var[NV-1] = -17.5 + 2.*random();*/

/*      simplx[*index].var[0] = -1400. + 160.*random();
        simplx[*index].var[1] = -3.5 + 0.5*random();
        simplx[*index].var[2] = -16000. + 2600.*random();
        simplx[*index].var[3] = -65. + 16.*random();
        simplx[*index].var[NV-1] = -15. + 1.*random();*/

/*      combines both duplex best answers
        simplx[*index].var[0] = -1400. + 570.*random();
        simplx[*index].var[1] = -3.5 + 1.5*random();
        simplx[*index].var[2] = -20500. + 7100.*random();
        simplx[*index].var[3] = -70. + 21.*random();
        simplx[*index].var[NV-1] = -17.5 + 3.5*random();*/
```

```
/*      simplx[*index].var[0] = -5400. + 8570.*random();
        simplx[*index].var[1] = -15.5 + 21.5*random();
        simplx[*index].var[2] = -50500. + 67100.*random();
        simplx[*index].var[3] = -170. + 221.*random();
        simplx[*index].var[NV-1] = -27.5 + 15.5*random();*/

/*      simplx[*index].var[0] = 600. + 100.*random();
        simplx[*index].var[1] = 0.0 + 6.*random();
        simplx[*index].var[2] = -42000. + 6000.*random();
        simplx[*index].var[3] = -170. + 70.*random();
        simplx[*index].var[NV-1] = -17. + 4.*random();*/

/*      simplx[*index].var[0] = -250. + 100.*random();
        simplx[*index].var[1] = -3.0 + 6.*random();
        simplx[*index].var[2] = -31000. + 6000.*random();
        simplx[*index].var[3] = -118. + 80.*random();
        simplx[*index].var[NV-1] = -17. + 4.*random();*/

/*      simplx[*index].var[0] = -4050. + 100.*random();
        simplx[*index].var[1] = -15.0 + 6.*random();
        simplx[*index].var[2] = 16600. + 6000.*random();
        simplx[*index].var[3] = 22. + 80.*random();
        simplx[*index].var[NV-1] = -16. + 4.*random();*/

/*      best triplex 2state answer*/
/*      simplx[*index].var[0] = -600. + 40.*random();
        simplx[*index].var[1] = -0.9 + 0.3*random();
        simplx[*index].var[2] = -4300. + 360.*random();
        simplx[*index].var[3] = -13. + 3.*random();
        simplx[*index].var[NV-1] = -11. + 1.*random();*/

        for(i=0;i<NV;i++)
                simplx[*index].var[i] = ext[2][i] + ext[3][i]*random();

/*      simplx[*index].var[1] = (simplx[*index].var[0]-4.0+(wt=random())*
                8.0)/295.0;
*/
/*      simplx[*index].var[3] = (simplx[*index].var[2]-4.0+(wt=random())*
                8.0)/295.0;
*/
#ifdef INITBUG
        if(dbgflg)
        fprintf(dbgout,"simplex point %d has values",*index);
        for(i=0;i<NV;i++)
```

```
                fprintf(dbgout,"%s%f",i%6? "   " : "\n\t\t",
                        simplx[*index].var[i]);
        fprintf(dbgout,"\n");
        fprintf(dbgout,"**************ENDSINIT**********************\n");
#endif

        return(*index);

}

/*-----------------------------------------------------*/

simcut()

{
int i,j;

        if(!(data = fopen("dk:simplx.mon","wmf")))
                {printf("can't open monitor file\n");
                return(FALSE);
                        }
        fprintf(data,"%s\n%d\n%d\n%d\n",simpar,nruns,monitor,next);
#ifdef SIMCUT
        printf("*******SIMCUT*************\n");
        printf("%s\n%d\n",simpar,nruns);
        for(i=0;i(NV+12;i++)printf(" %d ",rank[i]);
        printf("\n%f  %d\n\n",tolerance,circnt);
#endif
        for(j=0;j(NV+12;j++)
                {for(i=0;i(NV;i++)
                fprintf(data,"%s%f",(((i+1)%6)==0)? "\n" : " ",
                        simplx[j].var[i]);
                fprintf(data,"\n%f\n",simplx[j].sum);
                }
        for(i=0;i(NV+12;i++)fprintf(data," %d ",rank[i]);
        fprintf(data,"\n%f  %d\n\n",tolerance,circnt);
        fclose(data);
        printf("\n\nwhen you are through, please restart program SIMZIP\n");
        printf("on dk:, with dk:->dy1:\n\n\n\t\t\t\tthank you\n");
}
```

APPENDIX IV: IMPROVED METHODS OF ANALYSIS FOR CD DATA APPLIED TO SINGLE-STRAND STACKING[1]

CONTRIBUTION OF COAUTHORS

The paper contained in this appendix is presented as a sample application of the data analysis methods used in this thesis. The problem being investigated involved the relationship between CD spectra of synthetic polynucleotides and the degree of stacking of their bases in single strands. It was being conducted by Dr. Gary C. Causley in the laboratory of Professor W. Curtis Johnson, Jr. He performed the CD measurements, theoretical CD calculations, and theoretical development for decomposing observed CD spectra as Taylor series expansions of absorption spectra. My contribution to this work involved development of the computer-assisted directed search techniques used in solving the van't Hoff equation for various sets of melting data, including simplex and nonlinear least squares programming.

[1]Permission to include this manuscript was granted 5/16/86 by John Wiley and Sons, Inc. The manuscript was published in Biopolymers (1983), 22, p. 945-67.

Improved Methods of Analysis for CD Data Applied to Single-Strand Stacking

GARY C. CAUSLEY, PAUL W. STASKUS, and W. CURTIS JOHNSON, JR., *Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331*

Synopsis

CD spectra of ApA, poly(A), CpC, and poly(C), measured as a function of tempeature, are used to demonstrate improved methods of analysis for this type of data. Each of the four sets of temperature-dependent CD spectra are decomposed using singular value decomposition. This method demonstrates that there are only two independent parameters in each set of data. The two basis curves in the wavelength domain for each set of data are used to average the CD curves even though they are taken at different temperatures. Each of the four sets of temperature-dependent data, smoothed in this way, is decomposed in a Taylor series to separate degenerate interactions due to stacking from the intrinsic CD of the chromophores and nondegenerate interactions. Such a decomposition allows the comparison of corresponding interactions whose relative magnitudes are not obvious in the raw data. The CD due to degenerate interactions fits the van't Hoff equation well in all four cases and thermodynamic parameters are determined. A simplex routine is used to search the space before performing the least-squares fit of the data to the van't Hoff equation. This guarantees a global solution rather than a local solution. The first of the two basis curves in the temperature domain derived by singular value decomposition also fits the van't Hoff equation well and gives comparable thermodynamic parameters, although the basis curves derived in this manner have no *a priori* physical significance. Rather than comparing room temperature CD curves, we compare CD curves of the various compounds at their melting temperature where the populations of the stacked conformation are equivalent. We use the Cantor-Tinoco equation, which relates the CD of a polymer to the CD of the corresponding dimer, to show that the secondary structure of ApA is not the same as poly(A), and the secondary structure of CpC is not the same as poly(C). However, the two dimers have similar degenerate stacking interactions as do the two polymers. A sound theoretical basis is provided for the semiempirical Cantor-Tinoco equation.

INTRODUCTION

CD spectra have considerable information content with regards to secondary structure. In some biological polymers, the manifestation of a single structural feature can be the dominant spectral characteristic. One of the clearest examples of this phenomenon is found for the dimers and homopolymers of adenosine and cytidine, where stacking interactions cause the bulk of the CD intensity. As a result of this, these compounds have received a great deal of attention.[1–6] However, there are several types of stacking interactions manifest in the CD. The origins of these involve degenerate and nondegenerate perturbations of transition dipoles that are localized in the planes of the component bases. The different types of interactions

Biopolymers, Vol. 22, 945–967 (1983)

CCC 0006-3525/83/030945-23$03.30

complicate attempts to analyze the CD for secondary structure in these compounds.

Recognizing this problem, Tinoco has expressed the optical density (OD) and the CD spectra of oligomers as a Taylor series expansion about the OD of the monomeric unit[7] such that

$$\epsilon_p(\nu) = \bar{a}\epsilon_m(\nu) + \frac{\bar{b}\, d\epsilon_m(\nu)}{d\nu} \tag{1}$$

$$\Delta\epsilon_p(\nu) = a\epsilon_m(\nu) + \frac{b\, d\epsilon_m(\nu)}{d\nu} \tag{2}$$

where the $\bar{a}$, $\bar{b}$, a, and b coefficients are related to Coulomb interactions between the bases in a helical polymer. The theory predicts that the derivative terms result entirely from degenerate interactions of the stacked bases.[7-9] This term is referred to as the "conservative shape" in CD spectra.

Comparisons of this shape in CD curves are often made without considering the other variables that affect the data. For example, in the past, differences in maxima and minima[1] and changes in integrated intensities[2] have been used to monitor stacking interactions, but only in one case[7] have the contributions of degenerate and nondegenerate interactions been separated as prescribed by Eq. (2). Furthermore, in that work, the effects of temperature on stacking were not considered.

In this work we will do a complete analysis of the CD data for model oligomers, taking all variables that affect the shapes (including temperature) into consideration. Specifically, we will compare the CDs of dimers and polymers of adenosine and cytidine measured at temperatures between 0 and 90°C.

We are particularly interested in determining if the helical structure of the dimers is the same as that of the polymers. The semiempirical equation first used by Cantor and Tinoco[10]

$$\mathrm{CD}_{\mathrm{polymer}} = 2\,\mathrm{CD}_{\mathrm{dimer}} - \mathrm{CD}_{\mathrm{monomer}} \tag{3}$$

provides a simple relationship between our measured CD curves. We strengthen our confidence in this equation by showing that it has a sound theoretical basis. Then we use the relationship for the coefficients of the derivative terms in the Taylor series expansion. The equation is applied to comparable dimer and polymer coefficients taken at their respective melting temperatures where the populations of the stacked structures are the same. Our comparisons show the secondary structures of the dimers to be quite different from those of the corresponding polymers.

To achieve these goals, we first measure the CDs of adenylyl (3′ → 5′) adenosine (ApA), cytidylyl (3′ → 5′) cytidine (CpC), polyriboadenylic acid [poly(A)] and polyribocytidylic acid [poly(C)] over a wide temperature range. We then smooth the CDs and determine the number of orthogonal modes in them by utilizing the technique of singular value decomposi-

tion.[11-13] This technique provides an excellent way to smooth the data when multiple scanning is impractical. We decompose these smooth CD curves by fitting them to Taylor series expansions of adenosine 5′-monophosphate (AMP) and cytidine 5′-monophosphate (CMP) ODs. The coefficients determined in this analysis will be compared. We then fit the degenerate interaction terms to the van't Hoff equation.[14] The thermodynamic parameters thus obtained are comparable to literature values[15-17] and provide a method to determine the melting temperature of the oligomers.

Our work here represents a completed picture for these four compounds. The measurements are not new, but the method of analysis shows that without proper consideration of all the variables, erroneous conclusions about structure can be drawn from CD data.

EXPERIMENTAL

AMP, CMP, ApA, CpC, poly(A), and poly(C) were purchased from a variety of sources, including Sigma, Collaborative Research, and Miles Laboratories. ApA and CpC were used as the ammonium salts, because the free acids were found to contain high concentrations of the monomer. The purity of these compounds was monitored with high-pressure liquid chromatography (Varian model 5000). Polymers from Miles were used without further purification because they had high sedimentation coefficients. All samples were measured in 10 mM $Na_{3/2}H_{3/2}PO_4$, 10 mM in NaCl at a concentration of approximately $10^{-4}M$ nucleotide. Concentrations were determined from published extinction coefficients[18]: ϵ(AMP) = 15,100, ϵ(CMP) = 9200, ϵ(ApA) = 13,700, ϵ(CpC) = 8100, ϵ[poly(A)] = 10,400, and ϵ[poly(C)] = $6200M^{-1}$ cm^{-1}. The OD and CD spectra were measured on a Durrum Jasco model J-10 recording spectrograph, which was calibrated with (+)-10-camphorsulfonic acid (Aldrich) using $\Delta\epsilon$ = $2.37M^{-1}$ cm^{-1} at 290.5 nm. The recorded spectra were digitized at 1-nm resolution using a Hewlett-Packard model 9821A calculator equipped with a 9864A digitizing unit. Temperatures were regulated to 0.1°C with a thermoelectric temperature controller from Mechanical Design Co., Corvallis, OR.

Methods

Singular Value Decomposition

A set of n CD curves recorded as a function of temperature at m wavelengths is an $m \times n$ data matrix $\mathbf{Q}$. We wish to determine the number of significant orthogonal basis vectors in this data matrix, both in its range (wavelength) and its domain (temperature). Thus, Q is decomposed as

$$\mathbf{Q} = \mathbf{U}\Sigma\mathbf{V}^T \tag{4}$$

where U and V are unitary $m \times m$ and $n \times n$ matrices, respectively. The columns of U and V turn out to be what we seek, the orthonormal basis vectors in the range and domain of **Q**. Σ is an $m \times n$ matrix that has nonzero elements only on the "main diagonal" ($i = j$). These elements, σ_i, are the "singular values" of **Q**, that is, they are the positive square roots of the nonzero eigenvalues of $\mathbf{Q}^T\mathbf{Q}$, or equivalently $\mathbf{Q}\mathbf{Q}^T$.[11] As shown by Lloyd,[12] the product matrix $\Sigma\mathbf{V}^T$ is the matrix of least-square coefficients that will fit the basis vectors of U to **Q**. Conversely, $\mathbf{U}\Sigma$ is the matrix of least-square coefficients that will fit the basis vectors of $\mathbf{V}^T$ to **Q**. U, Σ, and V are determined from Jacobian diagonalization of

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{V}\Sigma^T\mathbf{U}^T\mathbf{U}\Sigma\mathbf{V}^T = \mathbf{V}(\Sigma^T\Sigma)\mathbf{V}^T \tag{5}$$

$$\mathbf{Q}\mathbf{Q}^T = \mathbf{U}\Sigma\mathbf{V}^T\mathbf{V}\Sigma^T\mathbf{U}^T = \mathbf{U}(\Sigma\Sigma^T)\mathbf{U}^T \tag{6}$$

The diagonal elements of $\Sigma^T\Sigma$ and $\Sigma\Sigma^T$ ($\sigma_1^2 \geq \sigma_2^2 \geq \sigma_k^2 > 0$) are the eigenvalues of $\mathbf{Q}^T\mathbf{Q}$ and $\mathbf{Q}\mathbf{Q}^T$ associated with the corresponding eigenvectors of V and U. Since σ_i^2 is the sum of squares of the intensities in **Q** that correspond to the ith eigenvector of U or V, the variance in **Q** that results from fitting β basis functions of m points is

$$\chi^2 = \frac{1}{m(k-\beta)} \sum_{i=\beta+1}^{k} \sigma_i^2 \tag{7}$$

where k is the number of nonzero elements of Σ. The standard deviation χ can be used to determine the number of significant orthogonal basis vectors in **Q**.

Taylor Series Fitting

Throughout this work, calculated and measured oligomer CD curves are fitted to Taylor series expansions of monomer absorption curves. The rationale for this treatment of the data has been given by Tinoco and Johnson.[8,9] This method takes into account the several phenomena that are observed in oligomer absorption and CD spectra. A summary of the major effects observed in absorption spectra includes[5,19]:

1. a splitting of the degeneracies in transition energies by Coulomb interactions among the monomers,
2. a shifting of the maximum of the resultant band because the component transitions that result from the splitting have different intensities, and
3. a change in absorption intensity resulting from mixing of the excited state with nondegenerate states of neighboring chromophores through Coulomb interactions.

Similarly, effects observed in the CD spectra of homopolymers include[5,19]:

1. Monomeric contributions, which are essentially the intrinsic CD of the chromophoric unit of the polymer. The shape of these CD bands is roughly the shape of the corresponding absorption.

2. Nondegenerate contributions, which result from mixing of the excited state with nondegenerate states of the neighboring chromophores through Coulomb interactions. This also results in CD bands with roughly the shape of the corresponding absorption.

3. Degenerate contributions, which result from a splitting of the degeneracies through Coulomb interactions among the monomers, leading to CD bands of various intensities and different signs for the resultant transitions. The total resulting CD must be zero, and the shape will resemble the derivative of the absorption band if the Coulomb interaction is small compared to the absorption bandwidth. It is this mechanism that represents the bulk of the CD observed here at lower temperatures.

Following Tinoco,[7] we begin with the shape of the monomer absorption band as a modified Gaussian[20] of the form

$$\epsilon_m(\nu) = 6.15 \times 10^{37}(\nu/\theta)D_a \exp[-(\nu - \nu_a)^2/\theta^2] \tag{8}$$

where D_a is the dipole strength of the monomeric unit and ν_a is the energy maximum of the monomer. θ is the half-width of the Gaussian with ν as the dependent variable. The CD for the monomer is then

$$\Delta\epsilon_m(\nu) = 2.46 \times 10^{38}(\nu/\theta)R_a \exp[-(\nu - \nu_a)^2/\theta^2] \tag{9}$$

where R_a is the rotational strength. For an oligomer of N monomeric units,

$$\epsilon_0(\nu) = 6.15 \times 10^{37} \frac{\nu}{N\theta} \sum_{k=1}^{N} D_k \exp\left[\frac{-(\nu - \nu_k)^2}{\theta^2}\right] \tag{10}$$

and

$$\Delta\epsilon_0(\nu) = 2.46 \times 10^{38} \frac{\nu}{N\theta} \sum_{k=1}^{N} R_k \exp\left[\frac{-(\nu - \nu_k)^2}{\theta^2}\right] \tag{11}$$

Remembering that Eqs. (10) and (11) for the absorption and CD of the oligomer are functions of $\nu - \nu_k$ and recognizing that the monomer curves are functionally dependent on $\nu - \nu_a$, it is clear that the oligomer curves may be expressed as a Taylor series expanded about $\nu - \nu_a$ such that with

$$f(\nu - \nu_k) = \exp[-(\nu - \nu_k)^2/\theta^2]$$

then,

$$f(\nu - \nu_k) = f(\nu - \nu_a) + (\nu_a - \nu_k)f'(\nu - \nu_a) + [(\nu_a - \nu_k)^2/2!]\, f''(\nu - \nu_a) + \cdots \tag{12}$$

Tinoco's original expansion about ν_a using ν_k as the dependent variable is only valid for polymers when ν_k can be considered a continuous variable. Instead, we have expanded about $\nu - \nu_a$, which is the same as applying the incremental form of the Taylor series expansion with increment $\nu_a - \nu_k$. This shows that the expansion [Eq. (12)] is actually valid for all oligomers.

Substitution of Eq. (12) into Eqs. (10) and (11) gives

$$\epsilon_0(\nu) = \frac{6.15 \times 10^{37}}{N\theta} \nu f(\nu - \nu_a) \sum_{k=1}^{N} D_k + \nu f'(\nu - \nu_a) \sum_{k=1}^{N} D_k(\nu_a - \nu_k) + \nu f''(\nu - \nu_a) \sum_{k=1}^{N} D_k \frac{(\nu_a - \nu_k)^2}{2!} \quad (13)$$

and

$$\Delta\epsilon_0(\nu) = \frac{2.46 \times 10^{38}}{N\theta} \nu f(\nu - \nu_a) \sum_{k=1}^{N} R_k + \nu f'(\nu - \nu_a) \sum_{k=1}^{N} R_k(\nu_a - \nu_k) + \nu f''(\nu - \nu_a) \sum_{k=1}^{N} R_k \frac{(\nu_a - \nu_k)^2}{2!} \quad (14)$$

Further,

$$\frac{d\epsilon_m(\nu)}{d\nu} \propto \frac{d[\nu f(\nu - \nu_a)]}{d\nu} = \nu f'(\nu - \nu_a) + f(\nu - \nu_a) \quad (15)$$

and

$$\frac{d^2\epsilon_m(\nu)}{d\nu^2} \propto \frac{d[\nu f'(\nu - \nu_a) + f(\nu - \nu_a)]}{d\nu} = \nu f''(\nu - \nu_a) + 2f'(\nu - \nu_a) \quad (16)$$

This gives Eq. (14) in terms of the extinction of the monomer and its derivatives such that

$$\Delta\epsilon_0(\nu) = \frac{4.0}{ND_a\theta} \sum_{k=1}^{N} R_k \epsilon_m(\nu) - \left[\sum_{k=1}^{N} R_k(\nu_k - \nu_a)\right]\left[\frac{d\epsilon_m(\nu)}{d\nu} - \frac{\epsilon_m(\nu)}{\nu}\right] + \left[\sum_{k=1}^{N} R_k \frac{(\nu_k - \nu_a)}{2!}\right]\left[\frac{d^2\epsilon_m(\nu)}{d\nu^2} - 2\frac{d\epsilon_m(\nu)/\nu}{d\nu}\right] + \cdots \quad (17)$$

or

$$\Delta\epsilon_0(\nu) = AG(\nu) + BG'(\nu) + CG''(\nu) \quad (18)$$

We will solve the integrated forms of Eq. (18) because this avoids the need to determine second derivatives, which are very noisy. The integrated form of Eq. (18) is

$$\int \Delta\epsilon_0(\nu)\, d\nu = A \int G(\nu)\, d\nu + B \int G'(\nu)\, d\nu + C \int G''(\nu)\, d\nu \quad (19)$$

Simpson one-third numerical integrals reduce the number of data points

per spectrum by factors of $2n - 1$; then, linear least-squares fits of this overdetermined system give the parameters of interest. Although the expansion is given to three terms, we are only interested in the first two, which are

$$A = \frac{4}{ND_a} \sum_{k=1}^{N} R_k \tag{20}$$

$$B = -\frac{4}{ND_a} \sum_{k=1}^{N} R_k(\nu_k - \nu_a) \tag{21}$$

Our coefficients differ from those of Tinoco[7] as the result of the way the variables in Eq. (19) are grouped. In his work, all terms containing $\epsilon_m(\nu)$ and $d\epsilon_m(\nu)/d\nu$ were grouped together. This results in Taylor series coefficients,

$$a = \frac{4}{ND_a} \sum_{k=1}^{N} R_k + \frac{1}{\nu} \sum_{k=1}^{N} R_k(\nu_k - \nu_a) \tag{22}$$

$$b = -\frac{4}{ND_a} \sum_{k=1}^{N} R_k(\nu_k - \nu_a) + \frac{2}{\nu} \sum_{k=1}^{N} R_k \frac{\nu_k - \nu_a}{2!} \tag{23}$$

which are more complicated than ours.

We solve for the simpler coefficients shown in Eqs. (20) and (21). The theoretical basis of these Taylor series coefficients has been discussed in the literature. Within the framework of exciton theory, B is related specifically to degenerate interactions and A to nondegenerate interactions and the intrinsic CD of the monomer itself. Since B results only from base stacking in the case of helical oligonucleotides, we will call B the "stacking coefficient." Furthermore, since the A and B coefficients are independent, each must satisfy Eq. (3) independently. We will make use of the relationship for the B coefficient,

$$B_{\text{polymer}} = 2B_{\text{dimer}} \tag{24}$$

where B for the monomer should be zero, since this term is due to stacking.

Finally, it is necessary to point out that the fitted curves, which are those from Eq. (19), are the sum of an unknown number of Taylor series terms. We determine that number of terms by calculating the average deviation in any given fit. We do this for calculated and measured CDs and have found that a minimum of three terms in the expansion are needed to reduce this deviation to an acceptable level (less than 10%). Here, the average deviation is defined as

$$\overline{AD} = \left[\sum_{i=1}^{m} \chi_i(\text{obs}) - \chi_i(\text{fit})\right] \Big/ \sum_{i=1}^{m} \chi_i(\text{obs}) \tag{25}$$

where m is the number of data points.

Melting Curves

Melting has been modeled as a two-state equilibrium between stacked and unstacked conformations in many studies of ribonucleic acid dimers

and polymers. To a good approximation, this has been shown to be an applicable model to the melting of stacking interactions of dimers. Optical data for polymers have also been fitted to the van't Hoff equation, but rationalization of scanning differential calorimetric data[17,21] (which yield different enthalpies and entropies) has led to the consideration of the Ising model[22,23] for a description of the melting process.[16,17] We consider both models (two-state and Ising) for our polymer melting curves.

First, we apply the two-state model to the temperature dependence of the stacking coefficients determined from Taylor series fitting, using the van't Hoff equation[14]:

$$B(T) = B_{un} + (B_{st} - B_{un})/[1 + \exp(\Delta H/RT - \Delta S/R)] \tag{26}$$

This equation was solved for the enthapy change (ΔH), entropy change (ΔS), unstacked coefficient (B_{un}), and stacked coefficient (B_{st}) by first searching the solution space using a simplex algorithm.[24] The simplex algorithm finds the area of the four-dimensional space which minimizes the sums of the squares of the differences between calculated and observed stacking coefficients, $B(T)$. A standard least-squares calculation[14] using the simplex solutions as a "best guess" solved for the parameters reported here.

Similarly, the Ising model is used to evaluate our polymer melting data. The equations that describe this model are given below in terms of the stacking parameter, B:

$$\begin{aligned} B(T) &= B_{un} + (B_{st} - B_{un})\alpha \\ \alpha &= 0.5 + 0.5(s-1)[(1-s)^2 + 4\sigma s]^{-1/2} \\ s &= \exp\left(\frac{-\Delta H}{RT} + \frac{\Delta S}{R}\right) \end{aligned} \tag{27}$$

We cannot evaluate σ (cooperativity factor) in this study However, we determine B_{st}, B_{un}, ΔH, and ΔS for the ranges of values of σ that have been published.[16,17,21] We show that for these ranges of σ, both models yield the same conclusions regarding structure.

Exciton Calculation

Following the procedure described by Bradley et al.,[25] a computer program was written to calculate the energies (E_k), dipole strengths (D_k), and rotatonal strengths (R_k) of a hypothetical dimer of identical monomers with one transition. In the nearest-neighbor approximation, the equations take the simple form

$$E_k = 2V_{12}\cos\frac{\pi k}{N+1} + V_{11}, \quad k = 1,2\ldots \tag{28}$$

$$D_k = \mu_\perp^2 S_1^2 + \mu_\perp^2 S_2^2 + \mu_\parallel^2 S_3^2 \tag{29}$$

and

$$R_k = \frac{\pi a}{c} \nu_0 \mu_\parallel \mu_t (S_3^2 - S_2^2 - S_1^2) + \frac{\pi z}{c} \nu_0 \mu_\perp^2 (S_2 S_4 - S_1 S_5) \qquad (30)$$

with

$$S_1 = \sum_{i=1}^{N} C_{ik} \cos \frac{2\pi}{p} i$$

$$S_2 = \sum_{i=1}^{N} C_{ik} \sin \frac{2\pi}{p} i$$

$$S_3 = \sum_{i=1}^{N} C_{ik}$$

$$S_4 = \sum_{i=1}^{N} iC_{ik} \cos \frac{2\pi}{p} i$$

$$S_5 = \sum_{i=1}^{N} iC_{ik} \sin \frac{2\pi}{p} i$$

The symbols are defined as follows: $\mu_\parallel$ and $\mu_\perp$ are the parallel and perpendicular components of the transition dipole μ, with respect to the helix axis; μ_r and μ_t are the radial and tangential components of μ; p is the number of monomers per turn of the helix; z is the rise per monomer along the helix axis; a is the distance of μ from the helix axis (see Fig. 1), and ν_0

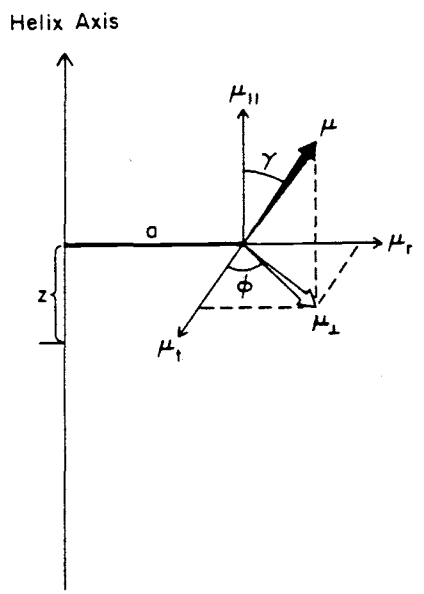

Fig. 1. Description of the transition dipole μ, $\mu_\parallel$ (parallel), $\mu_\perp$ (perpendicular), μ_t (tangential), and μ_r (radial) in terms of the angles ϕ and γ, where a is the distance from the center of the helix to the origin and z is the rise per base along the helix axis.

is the frequency of maximum absorbance for our hypothetical monomer. C_{ik} are the normalized mixing coefficients given by

$$\left(\frac{2}{N+1}\right)^{1/2} \sin \frac{\pi k_i}{N+1}$$

N is the number of monomers in our oligomer. V_{12} is the exciton coupling energy, and V_{11} is the state energy for the monomer.

Absorption and CD spectra for our hypothetical oligomers are generated using the Gaussian shapes for the bands given in Eqs. (10) and (11).

RESULTS AND DISCUSSION

Measured OD and CD

The melting of the dimers and polymers of adenosine and cytidine has been reported previously. The CD and ODs that we measure for this process are in good agreement with those reported in earlier studies.[1-7] For that reason we will not present all of the data. Instead, we present representative CD spectra for ApA, poly(A), CpC, and poly(C) in Figs. 2 and 3. In all cases, at least 16 temperature-dependent CD spectra were recorded. The new and significant part of this study is the analysis of the complete set of curves by the methods described above.

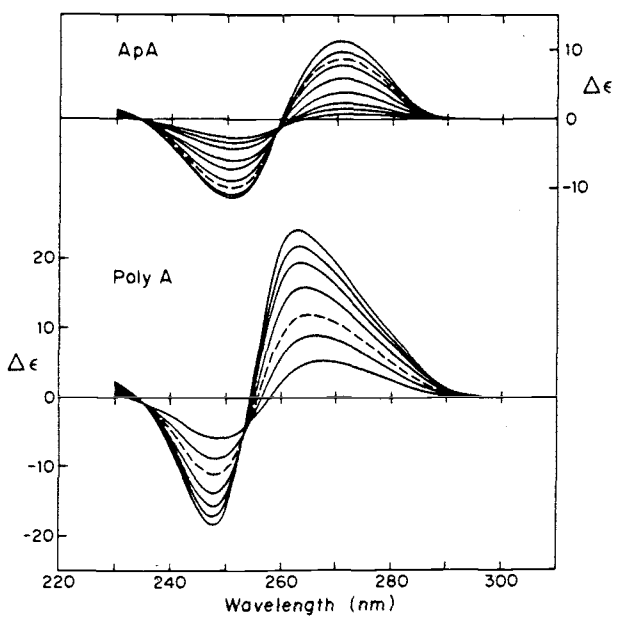

Fig. 2. Measured CD of ApA and poly(A). The dashed curves are the CDs at the melting points.

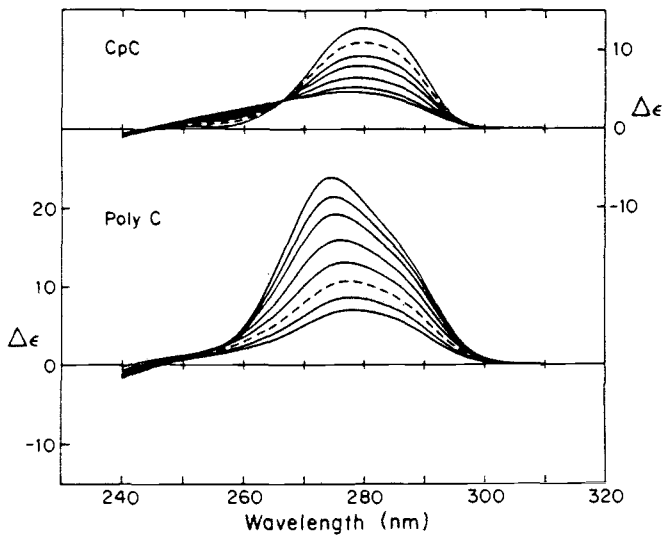

Fig. 3. Measured CD of CpC and poly(C). The dashed curves are the CDs at the melting points.

Singular Value Decomposition and Smoothing

The CDs were all analyzed to determine the number of significant singular values in each set of data. In all cases there were only two significant orthogonal bases. Thus, for each component we generate a U, Σ and V matrix of which only two vectors are necessary to reconstruct the experimental curves within experimental error. This is determined by the application of Eq. (7) to the values of Σ for each decomposition. The standard deviation that results from omission of all but β basis vectors in the reconstruction of the experimental CD curves for each compound is given in Table I for the first several values of β. It is clear from these results that two vectors will describe the data to an accuracy of 0.1–0.2 $\Delta\epsilon$ units. Accordingly, there are four important vectors (two temperature, two wavelength) for each compound. The two-dimensional representations of these vectors are given in Figs. 4 and 5. The temperature basis vectors are ΣV^T

TABLE I
Standard Deviation in Units of $\Delta\epsilon$ for Reconstruction of Experimental CD Data from β Basis Vectors Determined by Singular-Value Decomposition

Compound	$\beta = 1$	$\beta = 2$	$\beta = 3$	$\beta = 4$
ApA	0.47	0.07	0.06	0.05
poly(rA)	0.89	0.19	0.16	0.13
CpC	0.54	0.04	0.04	0.03
poly(rC)	0.44	0.12	0.07	0.04

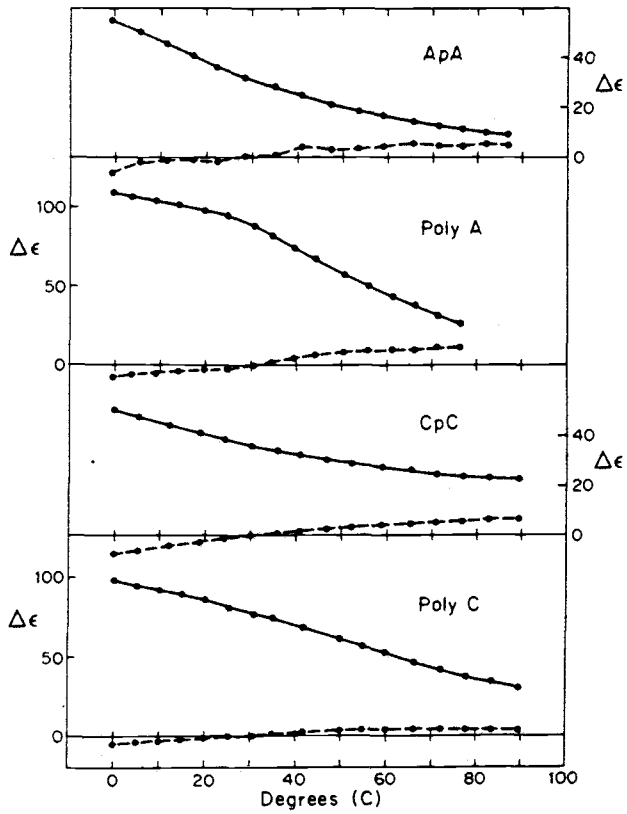

Fig. 4. The two most significant temperature basis vectors (ΣV^T) determined from singular-value decomposition of CD spectra.

and the wavelength basis vectors are $U\Sigma$, where only the first two bases are considered. We see that for each compound, the first temperature basis vector has a steep slope (Fig. 4). As we will show later, this basis vector fits the van't Hoff equation.

It is important to point out that *a priori* these vectors have no physical significance except that they are indicative of the number of independent orthogonal modes in the data sets. They must be related to physical reality by comparison to other data. However, the first temperature basis vectors do correspond to a wavelength basis vector (Fig. 5) that looks like the normally measured CD and is the least-squares[12] best average shape that changes as a function of temperature. In all cases, the second eigenvector is much less significant.

We will use the two most significant wavelength basis vectors to reconstruct the CD curves in any given melting series. This is an effective and efficient method for filtering the data of high-frequency noise.

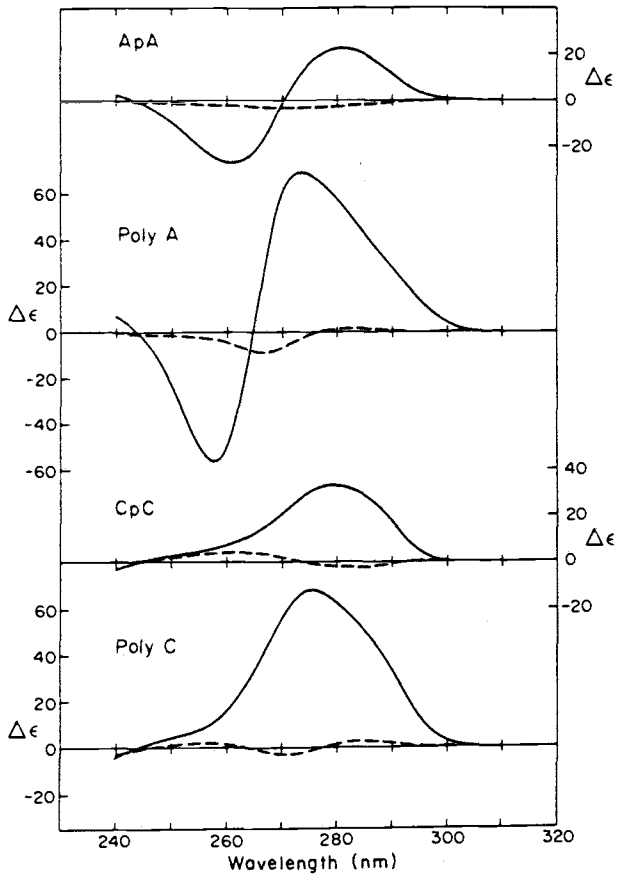

Fig. 5. The two most significant wavelength basis vectors ($\mathbf{U\Sigma}$) determined from singular-value decomposition of CD spectra.

Taylor Series Fits for *A* and *B*

The resultant smooth curves are decomposed into the Taylor series components as described above. The A and B coefficients determined for each compound at temperatures between 0 and 90°C are plotted in Fig. 6. The stacking coefficient, B, which corresponds to the derivative shape in the CD curves, is a dominant feature in all the curves at lower temperatures. This coefficient melts out, approaching zero as the temperature is increased, which in turn decreases the degenerate interactions.

The A coefficient, on the other hand, behaves quite differently over the temperature range studied. Here, the polymer values of A are much larger than the corresponding dimer values. The dimer values are nearly constant over the temperature range studied, and this fits well with the two-state model. The polymer A's have a steep slope, which does not fit the two-state model. In the case of poly(A), this coefficient even changes sign as the melt

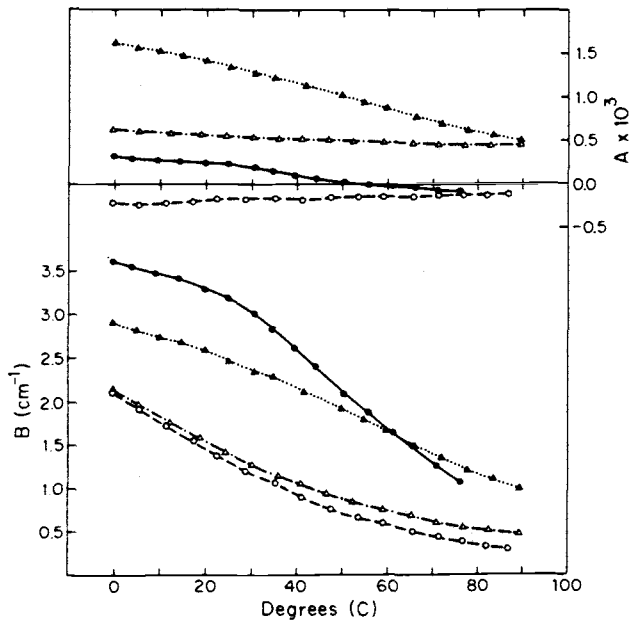

Fig. 6. Taylor series coefficients A and B for ApA (-○-), poly A (–●–), CpC (--△--), and poly(C) (··▲··) plotted as a function of temperature.

proceeds. This is not surprising, since the A term is a sum of several effects, only some of which are related to stacking. In particular, the change of monomer conformation as the polymer relaxes and the interaction of solvent with monomer are both known to change as a function of temperature.[26] These effects might not follow the two-state kinetics expected for stacking. This decomposition separates the degenerate part of the stacking effect so that it can be analyzed independently. If the entire spectrum is considered, then the other effects will have to be small if stacking is to be analyzed properly.

van't Hoff Equation

The Taylor series B *coefficients*, which are directly related to the stacking interactions, were found to fit the van't Hoff equation to an rms error of no more than 2%. The values for this stacking parameter at the low-temperature extrema and at the melting points, along with the pertinent thermodynamic constants, are given in Table II. A glance at these values shows clearly that B for the polymers is not twice B for the dimer, as predicted earlier. Indeed, B values for both poly(A) and poly(C) are almost the same as those for the corresponding dimers at the low-temperature extrema. The similarity of the B coefficients for the two dimers means that their stacking interactions are the same. This comparison

TABLE II
Compilation of Taylor Series Coefficients and Thermodynamic Parameters for Monomers, Dimers, and Polymers[a]

Compound	B_{st} (cm^{-1})	B_{mp} (cm^{-1})	A[b] ($\times 10^4$)	A_{mp} ($\times 10^4$)	$-\Delta H$ (kcal)	$-\Delta S$ (e.u.)
AMP[c]	0.0	—	−1.1	—	—	—
ApA	3.4	1.7	−2.2	−2.2	6.1	21.
poly(rA)	3.7	1.9	3.2	0.6	11.7	36.
CMP[c]	−0.1	—	4.7	—	—	—
CpC	3.4	1.8	6.2	5.9	5.9	21.
poly(rC)	3.2	1.4	16.2	7.5	6.5	19.

[a] ΔH and ΔS measured from the melting curves of B are in good agreement with other literature values.

[b] The value of A at the lowest temperature measured.

[c] Monomer A and B coefficients were invariant with temperature.

cannot be simply translated into conformation, because this one number depends on many variables. Nevertheless, it is consistent with the conformations of the two dimers being the same. This argument can also be made for the two polymers. As expected, the B values for all the compounds are quite close to zero at the high-temperature limit, where the degenerate interactions are minimal. The ΔH and ΔS values that result from the analysis are in good agreement with recently reported values for these parameters.[15–17]

The first temperature basis vectors from singular-value decomposition (Fig. 4) were also fit to the van't Hoff equation. These calculated parameters are given in Table III. The ΔH and ΔS values thus derived are comparable to those obtained with the stacking coefficient, B. This implies that the first wavelength eigenvectors (Fig. 5) are the bases that change in the CD curves as functions of temperature and that these changes are two state as described by the temperature basis vectors.

Ising Equation

The Taylor series B coefficients for poly(A) and poly(C) were found to give good fits to the Ising equation [Eq. (27)] over the range of σ's chosen.

TABLE III
Results of Fits of the First Temperature Eigenvector ($\Delta\epsilon$) from Singular-Value Decompositions to the van't Hoff Equation

Compound	$\sigma_1 v_1$ (st)	$\sigma_1 v_1$ (mp)	$\sigma_1 v_1$ (uns)	$-\Delta H$ (kcal)	$-\Delta S$ (e.u.)
ApA	87.	43.	−1.	6.0	21.
poly(A)	111.	58.	5.	11.9	37.
CpC	72.	44.	16.	5.6	20.
poly(C)	111.	52.	−7.	6.4	19.

TABLE IV
Compilation of Stacking Parameters and Thermodynamic Constants for Polymers Melting After the Ising Model

Compound	σ	B_{st} (cm^{-1})	B_{mp} (cm^{-1})	$-\Delta H$ (kcal)	$-\Delta S$ (e.u.)
Poly(rA)	0.4	3.8	2.0	7.2	22
	0.5	3.8	2.0	8.1	25
	0.6	3.7	2.0	8.9	27
Poly(rC)	0.8	3.2	1.5	5.8	17
	0.9	3.2	1.4	6.2	18
	1.0	3.2	1.4	6.5	19

For poly(A), cooperativities of 0.4–0.6 were used; and for poly(C), values between 0.8 and 1.0 were tested. Table IV contains a compilation of the results for these fits. The upper and lower limits of B are essentially insensitive to cooperativity. Thus, for either model our conclusions regarding structure as manifest in the stacking parameter are essentially the same.

Thermodynamic parameters are sensitive to cooperativity. Since the melting of poly(C) is thought to be fairly noncooperative (σ of 0.8–1.0), the values of ΔH and ΔS are similar for both models. However, the higher cooperativity for poly(A) (σ of 0.4–0.6) gives lower values of ΔH and ΔS for the Ising model than for the two-state model.

Semiempirical Relationship

Equation 3 was derived by Cantor and Tinoco[10] under a number of assumptions. Naturally, they assumed that the conformations of the polymer and corresponding dimer were the same, and they also made the reasonable assumption that only nearest-neighbor interactions need be considered. From these considerations it seemed only logical that the optical properties of an oligomer should be the sum of the optical properties of the component monomers plus the optical properties that result from nearest-neighbor (dimer) interactions. For the CD of a homopolymer, this relationship reduces to Eq. (3).

However, this line of reasoning is not so logical from a quantum-mechanical point of view. Under the nearest-neighbor assumption, the Hamiltonian for a polymer would be the sum of the Hamiltonians for the monomers plus the potential energies of the nearest-neighbor interactions, V_{12}. The polymer wave functions would be more complicated though, and involve the entire polymer even under the nearest-neighbor assumption. If we consider the simpler, defined problem of a homopolymer with one transition per monomer, there would be as many excited state wave functions as there are monomers. The corresponding transition energies would be split about the monomer transition energy over a range of $\pm 2V_{12}$.

In contrast, the dimer would have two transitions split by $\pm V_{12}$. Thus,

the many polymer transitions are crowded into an energy range that is only twice the splitting for the dimer. Each of the many closely spaced polymer transitions would have a CD band with a sign and intensity that can be determined from its corresponding wave function. It is not immediately obvious why the CD of a polymer, which results from the cancellation of a large number of closely spaced and overlapping bands, should have any simple relationship to the CD of a dimer that has two oppositely signed CD bands that do not cancel nearly so much because they are comparatively widely spaced. Nevertheless, many workers have shown with experimental data that the relationship works well.

In trying to rationalize the success of this relationship, we have discovered a sound theoretical basis. The Taylor series expansion [Eqs. (2) or (17)] provides the same simple shape for a CD curve regardless of the length of the oligomer. Thus, in this approximation we do not have to worry about relative splittings and the differing cancellations between dimers and polymers. Equation (26) from Johnson and Tinoco's[9] work shows that the contribution of the interactions to the first two Taylor series coefficients for the polymer are proportional to pairwise rotational strengths. In the nearest-neighbor approximation (assuming that the dimer and polymer have the same conformation), Eq. (3) is valid at least to the approximation that we can express the CD curves as the first two terms in a Taylor series.

Still, if we wish to use this equation to draw conclusions about conformation, we must be sure that we are not being misled by the nearest-neighbor assumption or the Taylor series approximation. Since we will be relying only on Eq. (24), we need only consider degenerate stacking interactions. In the next section we calculate CD curves in the nearest-neighbor approximation, but without the Taylor series expansion, and show that the relationship between the CD of the dimer and polymer remains the same. Further, we show that the pairwise rotational strength of more distant neighbors will not change our conclusions.

Calculated OD and CD

To further verify Eq. (3) as applied to our stacking coefficients [Eq. (24)], we have calculated the OD and CD for a homodimer and a homopolymer of 100 chromophoric units. We use a simple model with one transition in the nearest-neighbor approximation. With this model we can compare dimer to polymer and the effects of base tilt.

We are particularly interested in base-tilt effects, because our recent work in flow linear dichroism (to be published) and the results of electric linear dichroism[27,28] put the tilt angle of the bases in poly(A) and poly(C) at ca. 62° with respect to the helix axis.

For our model homopolymer we chose a structure like that of poly(A) as did Bradley et al.[25] We take eight bases per turn, a rise per base of 3.8 Å, a 3-Å radius to base center, a monomer half-width at half-height (θ) of

2300 cm^{-1}, a Coulomb interaction of 550 cm^{-1} (V_{12}), and a monomer transition dipole strength (D_a) of 16.6 D.[2] OD and CD curves for the monomer, the dimer, and the polymer were generated from Eqs. (10) and (11) using values of D_k and R_k from Eqs. (29) and (30).

The intensities vary with orientation of the transition dipole, but the shapes of the curves and the relationship between the dimer and the polymer are virtually independent of this parameter. Representative curves are shown in Fig. 7, where the transition dipole is oriented for a base tilt of 62° (γ) with no tangential component ($\phi = 90°$).

We now decompose the calculated CDs exactly as we did for the experimental curves, where a three-component Taylor series of the calculated monomer OD is fitted to the theoretically derived CDs. The theoretically determined B coefficient is a measure of the degenerate interactions in the fully stacked oligomer. These stacking parameters are collected in Table V.

We have also considered the effects of more distant neighbors. It was found that the second neighbor adds 44% of the CD intensity, the third neighbor adds about 8%, and the fourth neighbor subtracts about 8%. Thus, the theoretical prediction for the CD of our model polymer (as the result of degenerate interactions) is larger than twice the dimer CD when higher neighbors are considered. It would be correct to express the polymer B coefficient as a sum such that

$$B_{\text{polymer}} = (2B_{12} + 2B_{13} + \cdots + 2B_{1n}) \tag{31}$$

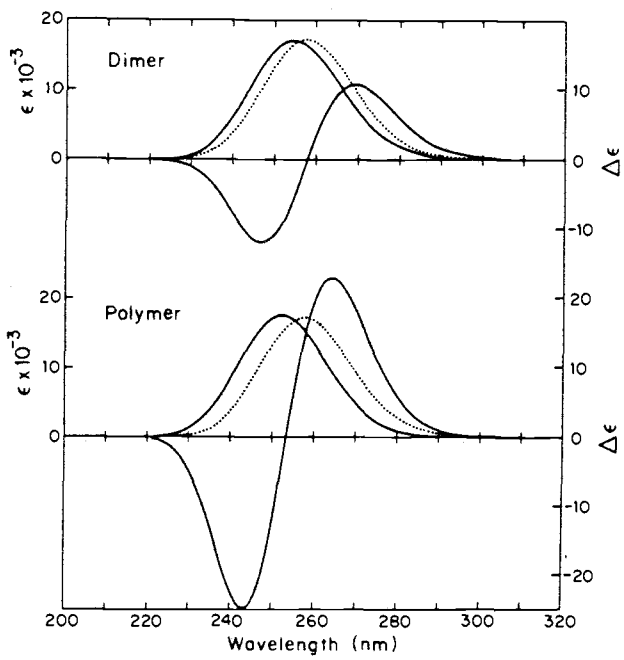

Fig. 7. Calculated OD and CD for a model monomer (···) and dimer and 100-mer (—).

TABLE V
Theoretical Coefficients for Degenerate Interactions (B) Obtained from Taylor Series Fits of Calculated Monomer OD to Calculated CDs of Dimer and 100-mer[a]

γ[b] (deg)	ϕ[b] (deg)	B (dimer)	B (100-mer)
62	90	1.8	3.4
62	45	2.1	3.9
62	0	2.7	5.0
90	NA	3.5	6.4

[a] Other parameters in the calculations remained constant (see Ref. 25).
[b] See Fig. 1.

Note that the various B values are positive for base angles less than 180°.

From Fig. 7 and Table IV the following observations can be made:

1. The blue shift for the OD of the polymer is twice that of the dimer.
2. The CD of the dimer has the crossover of its conservative shape at the wavelength of the *monomer* OD maximum.
3. The CD of the polymer is twice that of the dimer, but the crossover is moved to the maximum of the OD of the polymer.
4. The CD of both dimer and polymer are reduced for tilted bases. The greatest reduction is seen when there is no tangential component of the transition dipole.
5. In these fits, it is clear that for any given conformation,

$$B_{\text{polymer}} = 2B_{\text{dimer}} - B_{\text{monomer}}$$

in the nearest-neighbor approximation, where B_{monomer} is zero in our calculation—as expected.

6. The theoretically predicted B polymer coefficient would be even greater if higher neighbor effects were considered for a helix in the conformation studied here.
7. None of the theoretically predicted relations for the stacking coefficients hold for the experimental spectra (Table II). Thus, the conformation of each polymer must be different from the corresponding dimer.

Comparison of Dimer and Polymer CD

Workers often compare room-temperature CD spectra, but this is not reasonable for oligonucleotides that are in the middle of a conformational transition. Indeed, spectra should be compared only when the oligomers have equal fractions of the stacked forms. The best route would be to compare spectra when the oligomers are fully stacked, but such spectra cannot normally be measured. The next best route is to compare spectra at the melting temperature where $K_{eq} = 1$. For the oligomers considered here, the melting temperatures can be determined from the van't Hoff fits

discussed above. Furthermore, it is not really reasonable to compare the CD spectra themselves. Rather, the spectra should be decomposed and the effects of corresponding interactions compared, as discussed in the section on Taylor series fits.

Figure 8 shows the integral of the CD spectrum at the melting temperature and the integral of the corresponding Taylor series terms from Eq. (19):

$$A \int_{\nu_0}^{\nu_n} G(\nu)\, d\nu = A \int_{\nu_0}^{\nu_n} \epsilon_m(\nu)\, d\nu \tag{32}$$

$$B \int_{\nu_0}^{\nu_n} G'(\nu)\, d\nu = B \int_{\nu_0}^{\nu_n} \left[\frac{d\epsilon_m(\nu)}{d\nu} - \frac{\epsilon_m(\nu)}{\nu}\right] d\nu$$

$$= B\left[\epsilon_m(\nu)\Big|_{\nu_0}^{\nu_n} - \int_{\nu_0}^{\nu_n} \frac{\epsilon_m}{\nu}\, d\nu\right] \tag{33}$$

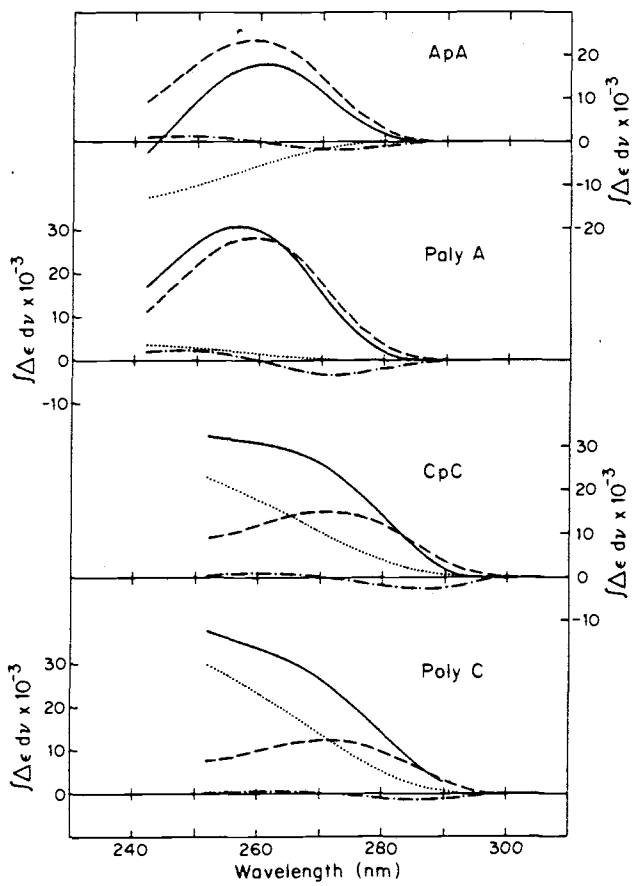

Fig. 8. Taylor series terms: $A \int G(\nu)\, d\nu$ (···), $B \int G'(\nu)\, d\nu$ (- - -), and $C \int G''(\nu)\, d\nu$ (-·-) and their sum (—) determined for each component at its melting point.

and

$$C\int_{\nu_0}^{\nu_n} G''(\nu)\,d\nu = C\int_{\nu_0}^{\nu_n}\left[\frac{d^2\epsilon_m(\nu)}{d\nu^2} - 2\frac{d(\epsilon_m(\nu)/\nu)}{d\nu}\right]d\nu$$

$$= C\left[\frac{d\epsilon_m(\nu)}{d\nu} - 2\frac{\epsilon_m(\nu)}{\nu}\right]\Bigg|_{\nu_0}^{\nu_n} \tag{34}$$

In all cases the contribution from stacking interactions [Eq. (33)] represents a large fraction of the total CD intensity and is similar in magnitude and shape for corresponding dimers and polymers. Note also that the A term is relatively small in oligomers of adenosine but is large in CpC and poly(C). This is because it is large in the monomer CD of cytidine (see Table II). The CD of the cytidine oligomers looks quite different from the CD observed for ApA and poly(A) because the degenerate terms are quite different in shape and the A term contributes significantly only to the cytidine oligomers.

Recently, Olsthoorn et al.[29-31] have reached many of the same conclusions for the oligomers of deoxyadenosine. As in this work on the ribooligomers, they observe a clear two-state behavior in the dimer but find relatively high values of ΔH and ΔS when the two-state model is applied to the polymer. They also conclude that the dimer has a different conformation from the polymer, but their documented reasons for this cannot be applied to our ribooligomers. In the deoxyoligomers the difference in conformation between the dimer and the polymer can be related through nmr measurements to the sugar conformation. The C2′-*endo* and C3′-*endo* deoxyribose conformations occur in a 70:30 ratio in the dimer. Apparently, the base–base interactions for the C2′,C2′-*endo* combination are minimal and contribute little intensity to the CD, whereas in the C2′,C3′-*endo* configuration, stacking interaction is considerable.

For ribooligomers, the sugars are all C3′-*endo*. With this arrangement the CDs of both dimer and polymer are intense. However, the CD of poly(A) is much less than one would predict from the dimer. If the polymer is primarily one type of secondary structure with essentially undetectable end effects, the data predict a different secondary structure for the dimer. This structure has an enhanced stacking interaction much like that observed in the deoxy dimer.

CONCLUSIONS

1. Raw CD curves cannot be used as a measure of stacking in helical polymers. Although the presence of a derivative-shaped curve in a CD spectrum indicates degenerate interactions, CD spectra must first be decomposed if the amount of stacking is to be assessed. Two methods for decomposition are discussed here.

2. Singular-value decomposition can be used to decompose CD spectra into orthogonal basis curves. Each of the four sets of temperature-de-

pendent CD spectra presented here [one each for ApA, poly(A), CpC and poly(C)] can be reproduced within experimental error by only two basis curves. The two basis curves can be used to smooth the data, effectively averaging all of the CD curves together, even though they were recorded at different temperatures. Each set of CD spectra can be decomposed into curves that are either a function of wavelength or a function of temperature. Although singular-value decomposition has the advantage of providing the number of independent parameters in a set of data (here two) and the advantage of providing a smoothing function for the data, it has the disadvantage that basis curves have no *a priori* physical significance.

3. If the CD spectra are decomposed in a Taylor series of the corresponding absorption spectrum, then the various terms do have a physical significance. The second term, which is a derivative, will be due only to degenerate interactions that result from stacking. The Taylor series decomposition allows workers to compare corresponding interactions between sets of CD data.

4. Since the derivative term in the Taylor series is related only to stacking, it can be used to study melting. This term fits the van't Hoff equation well for all four sets of data, and the derived thermodynamic parameters are given in Table II. The first basis curve in the temperature domain from singular-value decomposition also fits the van't Hoff equation well and gives comparable thermodynamic parameters (Table III).

5. The first Taylor series term, which corresponds to intrinsic CD of the chromophore and nondegenerate interactions, is much larger in the polymers than it is in the dimers. In the case of poly(A), it even changes sign as a function of temperature.

6. It is not reasonable to compare CD spectra of oligonucleotides taken at room temperature, because the samples may be at different points along their melting transitions. If spectra cannot be obtained for fully stacked samples, it is possible to compare spectra at the melting temperature where $K_{eq} = 1$. The melting temperature can be determined from a van't Hoff treatment of the decomposed data as described above.

7. We have compared stacking interactions of the dimer to the polymer. Neither for adenine nor for cystosine are the stacking interactions of the polymer twice the stacking interactions of the dimer as predicted by the Cantor-Tinoco equation. Thus, we conclude that the secondary structure of ApA is not the same as two adjacent bases in poly(A), nor is the secondary structure of CpC the same as two adjacent bases in poly(C). However, the degenerate stacking interactions of both fully stacked dimers are similar, as are these interactions for the two fully stacked polymers.

8. The semiempirical Cantor-Tinoco equation is shown to rest on a sound theoretical basis.

We would like to thank Drs. D. A. Goldman and C. K. Mathews for their help in determining the purity of ApA and CpC. This work was supported by NIH Grant GM-21479 from the Institute of General Medical Sciences and NSF Grant PCM 80-21210 from the Biophysics Program.

References

1. Brahms, J., Maurizot, J. C. & Michelson, A. M. (1967) *J. Mol. Biol.* **25**, 465–480.
2. Brahms, J., Michelson, A. M. & van Holde, K. E. (1966) *J. Mol. Biol.* **15**, 467–488.
3. Bush, C. A. & Brahms, J. (1967) *J. Chem. Phys.* **46**, 79–88.
4. Warshaw, M. M. & Cantor, C. R. (1970) *Biopolymers* **9**, 1079–1103.
5. Bloomfield, V. A., Crothers, D. M. & Tinoco, I., Jr. (1973) *Physical Chemistry of Nucleic Acids*, Harper & Row, New York, and references therein.
6. Ts'o, P. O. P. (1974) *Basic Principles in Nucleic Acid Chemistry*, Vol. 2, Ts'o, P. O. P., Ed., Academic Press, New York, p. 305, and references therein.
7. Tinoco, I. (1968) *J. Chim. Phys.* **65**, 715–731.
8. Johnson, W. C., Jr. & Tinoco, I., Jr. (1969) *Biopolymers* **8**, 715–731.
9. Johnson, W. C., Jr. & Tinoco, I., Jr. (1969) *Biopolymers* **7**, 727–749.
10. Cantor, C. R. & Tinoco, I., Jr. (1965) *J. Mol. Biol.* **13**, 65–77.
11. Noble, B. & Daniel, J. W. (1977) *Applied Linear Algebra*, Prentice-Hall, Englewood Cliffs, N.J., p. 323.
12. Lloyd, D. (1969) Ph.D. thesis, University of California, Berkeley.
13. Hennessey, J. P., Jr. & Johnson, W. C., Jr. (1981) *Biochemistry* **20**, 1085–1094.
14. Powell, J. T., Richards, E. G. & Gratzer, W. B. (1972) *Biopolymers* **11**, 235–250.
15. Frechet, D., Ehrlich, R. & Remy, P. (1979) *Nucleic Acids Res.* **7**, 1981–2001.
16. Dewey, T. G. & Turner, D. H. (1979) *Biochemistry* **18**, 5757–5762.
17. Freier, S. M., Hill, K. O., Dewey, T. G., Marky, L. A., Breslaner, K. J. & Turner, D. H. (1981) *Biochemistry* **20**, 1419–1426.
18. Richards, E. G. (1975) *Handbook of Biochemistry and Molecular Biology, Nucleic Acids*, Vol. 1, Fasman, G. D., Ed., C.R.C. Press, Cleveland, Ohio, p. 603.
19. Cantor, C. R. & Schimmel, P. R. (1980) *Biophysical Chemistry*, W. H. Freeman, San Francisco, p. 409, and references therein.
20. Lowry, T. M. (1964) *Optical Rotary Power*, Dover, New York, p. 396.
21. Suurkuusk, J., Alvarez, J., Freire, E. & Biltonen, R. (1977) *Biopolymers* **16**, 2641–2652.
22. Zimm, B. H. & Bragg, J. K. (1959) *J. Chem. Phys.* **31**, 526–535.
23. Applequist, J. (1963) *J. Chem. Phys.* **38**, 934–941.
24. Deming, S. N. & Morgan, S. L. (1973) *Anal. Chem.* **45**, 278A–283A.
25. Bradley, D. F., Tinoco, I., Jr. & Woody, R. W. (1963) *Biopolymers* **1**, 239–267.
26. Frechet, D., Ehrlich, R. & Remy, P. (1979) *Nucleic Acids Res.* **7**, 1965–1980.
27. Chen, H. H. & Charney, E. (1980) *Biopolymers* **19**, 2123–2132.
28. Charney, E. & Milstien, J. B. (1978) *Biopolymers* **17**, 1629–1655.
29. Olsthoorn, C. S. M., Bostelaar, L. J., van Boom, J. H. & Altona, C. (1980) *Eur. J. Biochem.* **112**, 95–110.
30. Olsthoorn, C. S. M., Bostelaar, L. J., De Rooij, J. F. M., van Boom, J. H. & Altona, C. (1981) *Eur. J. Biochem.* **115**, 309–321.
31. Olsthoorn, C. S. M., Haasnoot, C. A. G. & Altona, C. (1980) *Eur. J. Biochem.* **106**, 85–95.

Received May 6, 1982
Accepted August 12, 1982

APPENDIX V: COMPUTERIZED DATA ACQUISITION

The SUPERKIM (Lamar Instruments, Redondo Beach, CA) is a 6502 based microcomputer which can be dedicated to the task of collecting data from any source which provides the signal in the form of a DC voltage, preferably between -10 and +10 volts. The programming of the KIM, which was done directly in machine code, is truly NON interactive, so that operating the computer effectively requires some knowledge by the operator. There is an advantage to this, however, since as it now stands the computer can perform its function with a minimal set of peripherals such as a TTY or other terminal (ie it requires only a recorder for storage of data and programs on cassette tape). When linked to a more sophisticated computer such as a DEC 11/23, it should be a direct matter to create programming for operating the SUPERKIM from the higher computer in a more interactive manner, making many of the presently required manipulations invisible. Many of the ideas involved in the hardware and software development here were contributed by Dr. Walter A. Baase.

Hardware

The SUPERKIM uses 6522 chips (VIA's) for all I/O functions. These chips are described in the 6502 applications book by Rodnay Zaks (1979b), as well as in the SUPERKIM notebooks. The present computer configuration uses 2 6522's (U9 and U10), each of which has 2 timers, two I/O ports of 8 lines each, 4 control lines, and a wire OR connection to the 6502 interrupt request line (IRQ). Connections to the 2 I/O chips are as follows:

U10:

IRQ @pin 21 (jack4 pin4) → 6502 IRQ @pin 4 (jack1 pin26)
CA1 @pin 40 (jack5 pin1) → status line of A/D converter
CA2 @pin 39 (jack5 pin2) → R/C line of A/D converter
PA0 @pin 2 (jack5 pin3) → bit 4 of A/D converter
PA1 @pin 3 (jack5 pin4) → bit 5 of A/D converter
PA2 @pin 4 (jack5 pin5) → bit 6 of A/D converter
PA3 @pin 5 (jack5 pin6) → bit 7 of A/D converter
PA4 @pin 6 (jack5 pin7) → bit 8 (0) of A/D converter
PA5 @pin 7 (jack5 pin8) → bit 9 (1) of A/D converter
PA6 @pin 8 (jack5 pin9) → bit 10 (2) of A/D converter
PA7 @pin 9 (jack5 pin10) → bit 11 (3) of A/D converter

note that bits 0-3 of the A/D converter are wire-ORed with bits 8-11

CB2 @pin19 (jack5 pin20) → A0 of A/D
PB0 @pin10 (jack5 pin11) → enable line of multiplexer
PB1 @pin11 (jack5 pin12) → A0 of multiplexer
PB2 @pin12 (jack5 pin13) → A1 of multiplexer

```
PB3 @pin13  (jack5 pin14) → A2 of multiplexer
PB6 @pin16  (jack5 pin17) → PB7 @pin17  (jack5 pin18)
```

CA2 is used to tell the A/D converter when to convert, and CA1 reflects its status, telling the computer when it is finished. Port A is used to read in the converter output. CB2 tells the converter whether to report bits 0-3 or 4-11, and port B lines 0-3 are used to select one of the 8 lines (address) into the multiplexer to be transmitted to the A/D converter. PB6 is linked to PB7 because one of the timers on U10 will be counting pulses created by the other at the system rate of ~1MHz.

U9:

```
CA1 @pin40  (jack6 pin1)  → UART character xmitted
CA2 @pin39  (jack6 pin2)  → UART xmit character
PA0 @pin 2  (jack6 pin3)  → UART bit 0
PA1 @pin 3  (jack6 pin4)  → UART bit 1
PA2 @pin 4  (jack6 pin5)  → UART bit 2
PA3 @pin 5  (jack6 pin6)  → UART bit 3
PA4 @pin 6  (jack6 pin7)  → UART bit 4
PA5 @pin 7  (jack6 pin8)  → UART bit 5
PA6 @pin 8  (jack6 pin9)  → UART bit 6
PA7 @pin 9  (jack6 pin10) → UART bit 7
PB7 @pin17  (jack6 pin18) → UART xmit timing
PB6 @pin16  (jack6 pin17) → 60Hz pulse train
IRQ @pin21  (jack5 pin21) → 6502 IRQ @pin4  (jack1 pin26)
```

The SUPERKIM operates its A/D converter (Analog Devices AD 574) at a frequency of approximately 2 KHz. The pair of timers on U10 count down between conversions at the 1 MHz system frequency. Depending upon the data collection rate, a certain number of A/D conversions are averaged, the result is converted to binary coded decimal (BCD), and output to memory as well as a terminal through a Universal Asynchronous Receiver Transmitter (UART). One timer on U9 provides timing pulses for UART transmission at 110 baud. There is a 60 Hz clock, composed of a 74121 one-shot chip with appropriate time constant, which uses the house AC signal after passage through a signal diode and resistor. The output of the one-shot is connected to pin 17 of 6522 U9 (PB6). U9's timer #2 counts the pulses to keep track of real time data collection intervals.

Circuit boards

These notes on making circuit boards come from the laboratory of Dr. James Ingle, in the chemistry department at OSU. Rub both sides of circuit board material (copper coated) vigorously with abrasive cleanser, rinse with distilled water and dry in an oven (placed to drain well). Spray surfaces to coat with photoresist, with inert

carrier gas at 20 pounds/square inch. Oven dry boards about 20 minutes. Place circuit pattern on board and make sure it is flush. Expose board with pattern for about 10 minutes to short wavelength visible light, then place in $FeCl_3$ bath with heating to remove unwanted conductor. I think the circuit pattern can be as simple as a "Sharpie" drawing on clear plastic.

Software

Without a sophisticated host computer, certain registers of the SUPERKIM must be filled with scan parameters by hand before the data are collected. These registers and the hexadecimal values to be stored in them are presented in Table AV.1. Instructions for manipulating KIM registers, as well as a description of hexadecimal values and instructions for loading or reading KIM memory blocks via magnetic tape or terminal, can be found in the KIM-1 User Manual (1976), MOS Technology Inc., Norristown, PA.

In addition to the registers in Table AV.1, certain other locations must be attended to before data collection, as follows:

register address	value
0017	multiplexer address of primary data line
0018	multiplexer address of secondary data line
001C	number of data points in scan (lo, in HEX)
001D	number of data points in scan (hi, in HEX)
0022	address for storage of primary line data (lo)
0023	address for storage of primary line data (hi)
0024	address for storage of secondary line data (lo)
0025	address for storage of secondary line data (hi)
0026	# of 60Hz cycles in delay after external trigger (lo)
0027	# of 60Hz cycles in delay after external trigger (hi)
00F1	arithmetic flag (should be 00)
17FA	non-maskable interrupt address lo (should be 00)
17FB	non-maskable interrupt address hi (should be 1C)

During program execution, data from the primary line only is sent to a terminal or TTY through the UART in a compressed format with two nybbles of information per byte. It can be punched on paper tape or collected by the host computer. The secondary data can be collected only after the data scan has been completed.

This documentation describes the program written for the SUPERKIM to record voltage data on a time base. Since the program was initially written in machine code (i.e. I did the assembling), the mnemonic code PROBABLY WILL NOT assemble without error if isolated in a source file for the 6502 assembler. It could be pretty close. Meanings of machine and mnemonic codes can be found in the programming manual by Zaks (1979a). Description of the 6522 chip and its control is in the applications manual (Zaks, 1979b). Address labels are in small print and program comments follow the ';' character. The ellipsis (...) in the code represents unused memory registers or no operation (NOP's).

```
machine          mnemonic
address  code    code                  comment
(hex)    (hex)

0200   A9 BF     LDA #$BF     ;put "BF" in U10's DDRB, so that bit
       8D 0213   STA  $1302   ;#6 is the only input line on port B

       A5 20     LDA  $00     ;load U9's T2 low latch for data rate
       8D 0812   STA  $1208   ;60 Hz countdown.
                              ;20 and 21 contain a HEX value which
                              ;corresponds to the number of 60Hz
                              ;pulses (taken from house AC) which
                              ;will occur for each datum time
                              ;interval

       A9 E0     LDA #$EO     ;put "EO" in U9's ACR (T1 makes
       8D 0B12   STA  $120B   ;continuous square wave pulses for
                              ;the UART on PB7, T2 counts pulses
                              ;on PB6 from the house AC

       A5 21     LDA  $0021   ;load U9's T2 high counter (60 Hz data
       8D 0912   STA  $1209   ;rate) and start counting pulses on
                              ;PB6

       A9 60     LDA #$60     ;put "60" in U10's ACR (T1 makes
       8D 0B13   STA  $130B   ;continuous interrupts and T2 counts
                              ;pulses on PB6)

0219   8D FE17   STA  $17FE   ;"60" also happens to be low address
                              ;for interrupt service routine

       A9 CE     LDA #$CE     ;put "CE" in U10's PCR to hold CA2
                              ;high
       8D 0C13   STA  $130C   ;and hold CB2 low

       A9 7F     LDA #$7F     ;disable all of U10's interrupts
       8D 0E13   STA  $130E

       A9 03     LDA #03      ;"03" is interrupt routine high
       8D FF17   STA  $17FF   ;address.
```

```
        A5 1E       LDA  $001E   ;load T1's low latch with low part of
        8D 0613     STA  $1306   ;system clock pulses per A/D conver-
                                 ;sion.
                                 ;1E and 1F contain a HEX value which
                                 ;corresponds to the number of computer
                                 ;crystal cycles (~1MHz) which will
                                 ;elapse between successive conversions
                                 ;of the A/D converter

0230    A9 C0       LDA #$C0     ;enable T1 interrupt on U10.  this is
        8D 0E13     STA  $130E   ;the only interrupt enabled, which
                                 ;tells the computer that it's time
                                 ;for another A/D conversion

        A9 1C       LDA #$1C     ;high address for non-maskable
        8D FB17     STA  $17FB   ;interrupt (makes reset and single-
                                 ;step mode work)

        A2 04       LDX #04      ;"04" is offset for first data line
                                 ;data storage, put it in X register

        A9 00       LDA #00      ;zero accumulator and initialize
        85 31       STA  $0031   ;various intermediate result registers
        85 0C       STA  $000C   ;0C-0F will contain the sum of A/D
        85 0D       STA  $000D   ;conversions
        85 0E       STA  $000E
        85 0F       STA  $000F
0248    85 10       STA  $0010
        85 11       STA  $0011
        85 12       STA  $0012
        85 13       STA  $0013
        85 2B       STA  $002B   ;2B and 2C used for TTY output
        85 2C       STA  $002C   ;checksum
        85 30       STA  $0030
        8D FA17     STA  $17FA   ;low address for non-maskable
        85 04       STA  $0004   ;interrupt.

        A5 1A       LDA  $001A   ;registers 1A and 1B contain the
        85 15       STA  $0015   ;number of A/D conversions averaged per
        A5 1B       LDA  $001B   ;datum.  transfer values to countdown
0261    85 16       STA  $0016   ;registers 15 and 16

        A9 02       LDA #02      ;intialize 2A at "02".  2A is used as
        85 2A       STA  $002A   ;counter register.  it typically starts
                                 ;from the HEX value "D0" and counts up.
;"D0" in binary is 1101 0000.  Since the leftmost bit is 1, the number
;is considered negative by the 6502 processor, corresponding to -48
;decimal.  There are 48 ascii characters in each line of TTY output.
;while 2A is negative, it prompts the sending of data values to the
;terminal.  When passing through zero, however, it signals the
;transmission of other characters as follows:
;       0       send high part of checksum for data line on terminal
;       1       send rest of checksum
```

```
;       2       send ascii <carriage return>
;       2       send ascii <line feed>
;       4       send ascii "+" as a flag for a new line of data
;       5       reinit. 2A at "D0", and checksum registers at "00"

        A9 00       LDA #00
        85 28       STA  $0028    ;low address of data block to be sent
                                  ;to terminal is always started at 00
        85 2E       STA  $002E    ;2E contains this same value as an
                                  ;offset from a base address for data
                                  ;storage blocks.  why not let both of
                                  ;these values start as the contents
                                  ;of 22?

        EA
        EA                        ;no op's

        A5 23       LDA  $0023    ;high address for data line 1 data
        85 29       STA  $0029    ;storage block. 28 and 29 point
                                  ;to THIS data block for output to
                                  ;TTY during program execution

        C6 29       DEC  $0029    ;decrement this address WHY?

        A9 FF       LDA #$FF      ;make U9's ORA lines all outputs
        8D 0312     STA  $1203

027A    A9 08       LDA #08       ;make U9's ORB lines all inputs
        8D 0212     STA  $1202    ;except but 3

        A9 7F       LDA #$7F      ;disable all of U9's interrupts
        8D 0E12     STA  $120E

        A9 19       LDA #$19
        8D 0412     STA  $1204    ;U9's T1 low latch

        A9 01       LDA #01
        85 09       STA  $0009

        8D 0512     STA  $1205    ;U9's T1 high counter.  T1 provides
                                  ;timing for UART to TTY at 110 BAUD

0290    A9 0F       LDA #$0F      ;put "0F" in U9's PCR to hold CA2 high
        8D 0C12     STA  $120C    ;and have CA1 interrupt on low-high
                                  ;transition for UART signals

        A5 1F       LDA  $001F    ;load U10's T1 high counter with time
        8D 0513     STA  $1305    ;per A/D conversion  ; start interrupt
                                  ;train

        20 C003     JSR ad        ;jump to A/D conversion subroutine

        88          DEY           ;decrement Y register
```

```
        4C FA02     JMP  data     ;send byte to UART
chk     A5 09       LDA  $0009    ;check register 09 contents, and branch
        F0 FC       BEQ  chk      ;if zero (ie no bytes to type)
                                  ;09 contains the number of data bytes
                                  ;to be sent to the terminal during
                                  ;program execution, in HEX

        B1 28       LDA ($0028),Y;get next data point stored

        18          CLC           ;clear carry bit and set decimal flag
02A8    F8          SED

        65 2B       ADC  $002B    ;add last data point into register 2B
        85 2B       STA  $002B    ;and any carry bit into 2C this is for
        A5 2C       LDA  $002C    ;checksum of data going to terminal?
02AF    69 00       ADC #00
        85 2C       STA  $002C

        D8          CLD           ;clear decimal flag

        B1 28       LDA ($0028),Y;get last data point and send to
        4C C802     JMP  $02C8    ;UART

        ...EA...

************************************
following section sends characters to UART for output
************************************

02C8    8D 0112     STA  $1201    ;put it in U9's ORA (goes to UART)

        A9 0D       LDA #$0D      ;put "0D" in U9's PCR to hold CA2 low
        8D 0C12     STA  $120C    ;and have CA1 and CB1 trigger on a
                                  ;high→low transition

        A9 0F       LDA #$0F      ;move CA2 high
        8D 0C12     STA  $120C

cal     AD 0D12     LDA  $120D    ;check CA1 flag in U9's IFR and
02D8    29 02       AND #02       ;branch if not set (ie UART not done)
        F0 F9       BEQ cal

        E6 2A       INC  $002A    ;UART done.  increment counter for
                                  ;# of data/line on terminal

        F0 15       BEQ nwlin     ;branch if zero for newline cr
        10 18       BPL data      ;branch if plus for character output

02E2    C6 09       DEC  $0009    ;decrement counter of bytes to type
        D0 04       BNE more      ;branch if not 0 (ie more to do)

        A5 3D       LDA  $003D    ;check # data points high counter
        30 08       BMI finis     ;branch if negative, last point
```

```
more    C8          INY           ;increment Y register index for data
        D0 02       BNE y1        ;branch if not zero, since not a new
                                  ;page of data and high storage address
                                  ;need not be incremented

        E6 29       INC  $0029    ;increment data storage high address

y1      4C A102     JMP chk       ;go to wait loop

finis   4C 3503     JMP  $0335    ;done, clean up

nwlin   A5 2C       LDA  $002C    ;get part of checksum
        4C C802     JMP  $02C8    ;send to UART

data    A5 2A       LDA  $002A    ;get 002A
        18          CLC           ;clear carry
        6A          ROR           ;rotate right
        D0 05       BNE not1      ;branch if not 0 now (ie 002A not = 1)

        A5 2B       LDA  $002B    ;otherwise, get 002B
        4C C802     JMP  $02C8    ;and send out on UART

not1    6A          ROR           ;rotate 002A right again
        F0 08       BEQ is2       ;branch if now zero (ie 002A is 2)
0308    30 0B       BMI gr2       ;branch if minus (ie 002A > 2)

        6A          ROR           ;rotate 002A right again
        F0 23       BEQ is4       ;branch if zero now (ie 002A is 4)
        4C CF05     JMP  $05CF    ;error condition

is2     A9 8D       LDA #$8D      ;get "8D" (ascii cr)
        4C C802     JMP  $02C8    ;and send it to UART

gr2     6A          ROR           ;rotate 002A right again
        30 06       BMI is37      ;branch if minus (002A is 3,7,...)
        6A          ROR           ;rotate 002A right again
        30 08       BMI is5       ;branch if minus (002A is 5)
        4C D605     JMP  $05D6    ;error condition

is37    A9 0A       LDA #$0A      ;get "0A" (ascii lf)
0320    4C C802     JMP  $02C8    ;and send it to UART

is5     A9 D0       LDA #$D0      ;get "D0" char. per line on terminal
        85 2A       STA  $002A    ;and reinit. 002A
        A9 00       LDA #00       ;reinit. 002B and 002C
        85 2B       STA  $002B
        85 2C       STA  $002C
        4C E202     JMP  $02E2    ;check for more bytes to send to UART

        A9 2B is4   LDA #$2B      ;get "2B" (ascii +) to start next
        4C C802     JMP  $02C8    ;terminal line and send it to UART
```

```
*****************************************************
this section is for cleanup before program termination
*****************************************

0335    A9 40       LDA #$40      ;put "40" in U10's IER to disable
        8D 0E13     STA  $130E    ;T1 interrupts

033A    A5 42       LDA  $0042    ;get low address of last datum stored
        8D F717     STA  $17F7    ;and put it in 17F7 for sending to
        A5 43       LDA  $0043    ;TTY or 11/2. likewise for high address
        8D F817     STA  $17F8    ;to 17F8

        A9 00       LDA #00       ;restore interrupt routine address
        8D FE17     STA  $17FE
        A9 1C       LDA #$1C
        8D FF17     STA  $17FF
        4C 9005     JMP  $0590    ;finish cleanup or restart scan (model
E)

*****************************************************
here is interrupt service routine
*****************************************************

0360    48          PHA           ;put accum. on stack
        98          TYA           ;save Y register too
        48          PHA
        D8          CLD           ;clear decimal flag

        AD 0D13     LDA  $130D    ;get U10's IFR
        10 09       BPL not10     ;branch if plus (not U10 interrupt)
0369    2A          ROL           ;roll 130D left
        10 06       BPL not10     ;branch if plus (not the T1 interrupt)
        AD 0413     LDA  $1304    ;to clear U10's T1 flag

        20 BB03     JSR  $03BB    ;do A/D conversion

not10   68          PLA           ;restore Y register
        A8          TAY
        68          PLA           ;and accum.
        58          CLI           ;clear interrupt flag
        40          RTI           ;return from interrupt

        ..............

*************************************************************
this section is used to create a delay after external triggering by
such as the model E UV scanner
*************************************************************

038B    A5 26       LDA  $0026    ;load U9's T2 low latch and high
        8D 0812     STA  $1208    ;counter with pulse count (for delay)
        A5 27       LDA  $0027    ;after trigger for start of data
        8D 0912     STA  $1209    ;collection (model E)
```

```
nottim AD 0D12    LDA  $120D   ;check U9's IFR T2 flag
0398   29 20      AND #$20
       F0 F9      BEQ nottim   ;and branch if not set

       AD 0812    LDA  $1208   ;U9's T2 timed out, clear its flag

************************************************
this is start of program for typical time base collection, manually
starting computer in synch with apparatus being monitored
************************************************

039F   A5 1C      LDA  $001C   ;copy 1C to 3C, 1C and 1D contain the
       85 3C      STA  $003C   ;number of points in the spectral scan
                               ;3C and 3D will serve as decrementing
       A5 1D      LDA  $001D   ;copies
       85 3D      STA  $003D   ;copy 1D to 3D

       A5 22      LDA  $0022   ;copy 22 to 42 ; 22-25 contain the low
       85 42      STA  $0042   ;and high addresses for storage blocks
                               ;of the data from 2 lines
       A5 23      LDA  $0023   ;same for 23,24,25
       85 43      STA  $0043   ;42-45 will be incremented as data is
       A5 24      LDA  $0024   ;stored to keep track of the last
03B1   85 44      STA  $0044   ;datum's address
       A5 25      LDA  $0025
       85 45      STA  $0045

03B7   4C 0002    JMP  $0200   ;continue initialization at 200

************************************************

03BB   A5 3D      LDA  $003D   ;check 3D and branch if there are
       10 01      BPL ad       ;data points left to record

       60         RTS          ;else return from subroutine

ad     A4 2E  ad  LDY  $002E   ;Y register gets 2E.  2E contains the
                               ;offset from a base address for data
                               ;storage in a block

       8A         TXA          ;get X register value (tells line)
       F0 1E      BEQ secnd    ;and branch if 0 (2nd line)
                               ;00 is line 2, 04 is line 1

       ...EA...   NOP          ;no ops

       A2 00      LDX #00      ;next pass will be 2nd line input

       A5 17      LDA  $0017   ;get multiplexer address for 1st line
       8D 0013    STA  $1300   ;and put in U10's ORB

       ...EA...   NOP          ;no ops
```

```
        A5 43       LDA  $0043   ;low address for line 1 data storage
        85 05       STA  $0005   ;passed to 05

        ...EA...    NOP          ;no ops

        4C 0004     JMP  $0400   ;go to 400 to skip 2nd line manip.
03E0    4C DD05     JMP  $05DD   ;error condition

secnd   A5 18       LDA  $0018   ;2nd line multiplex address goes to
        8D 0013     STA  $1300   ;U10's ORB

        ...EA...    NOP          ;no ops

        A5 45       LDA  $0045   ;low address for line 2 data storage
        85 05       STA  $0005   ;passed to 05

        A2 04       LDX #04      ;next pass will be 1st line input

        A5 15       LDA  $0015   ;check low and high parts of countdown
        D0 08       BNE nohi     ;for A/D conversions per datum, and
        A5 16       LDA  $0016   ;if positive (not program termination)
03F8    D0 02       BNE both     ;increment them since this is 2nd line

        88          DEY          ;decrement Y register why?
        88          DEY          ;only if last A/D for datum!

both    E6 16       INC 0016
nohi    E6 15       INC 0015

0400    A9 CC       LDA #$CC     ;put "CC" in U10's PCR to hold CA2 and
        8D 0C13     STA  $130C   ;CB2 low  this triggers A/D to convert

        AD 0413     LDA  $1304   ;clear U10's T1 interrupt flag

        ...EA...    NOP          ;no ops

nocal   AD 0D13     LDA  $130D   ;check U10's IFR CA1 flag
        29 02       AND  #02
        F0 F9       BEQ nocal    ;and branch if not set

0411    A9 EE       LDA #$EE     ;CA1 tripped, converter done, so flip
        8D 0C13     STA  $130C   ;U10's CA2 and CB2 high

        AD 0113     LDA  $1301   ;read data at U10's ORA (this is low
                                 ;nybble)

        18          CLC          ;clear carry

        75 0C       ADC  $000C,X ;add data to 0C+X
        95 0C       STA  $000C,X

        A9 CE       LDA #$CE     ;now flip U10's CB2 low (via PCR)
        8D 0C13     STA  $130C
```

```
        AD 0113     LDA  $1301    ;and read high byte of data at U10's
                                  ;ORA

        75 0D       ADC  $000D,X ;add data to 0D+X
0428    95 0D       STA  $000D,X

        A9 00       LDA #00       ;add carry overflow into 0E+X and
        75 0E       ADC  $000E,X ;0F+X
        95 0E       STA  $000E,X
        A9 00       LDA #00
        75 0F       ADC  $000F,X
        95 0F       STA  $000F,X

        B0 A8       BCS           ;(-88 steps)data overflow condition

        C6 15       DEC  $0015    ;decrement counters for A/D conversions
        D0 04       BNE cntlo     ;per datum
        C6 16       DEC  $0016
        F0 42       BEQ last      ;branch if last conversion needed for
                                  ;this datum

cntlo   60          RTS           ;return from subroutine
        EA          NOP

0442    C9 08       CMP #08       ;is it 8? (2K)
        F0 12       BEQ is8
        C9 04       CMP #04       ;is it 4? (1K)
        F0 0D       BEQ is4
        C9 02       CMP #02       ;is it 2? (.5K)
        F0 08       BEQ is2
        C9 01       CMP #01       ;is it 1? (.25K)
        F0 03       BEQ is1
        4C E405     JMP  $05E4    ;error condition
is1     C8          INY           ;add up # of rolls to read data sum
is2     C8          INY           ;in Y register
is4     C8          INY
is8     C8          INY

roll    18          CLC           ;clear carry
        36 0D       ROL  $000D,X ;roll data collection registers
        36 0E       ROL  $000E,X ;to left to prepare to read average
        36 0F       ROL  $000F,X
        88          DEY           ;count down rolls in Y register
        D0 F6       BNE roll      ;until done

        4C CB04     JMP  $04CB    ;do BCD conversion

        A9 08       LDA #08
        85 40       STA  $0040
        4C 4F1C     JMP  $1C4F    ;detect error condition
        A9 09       LDA #09
        85 40       STA  $0040    ;40 used to hold
0471    4C 4F1C     JMP  $1C4F    ;numbers and detect various error
```

```
        A9 0A       LDA #$0A      ;conditions before return to
        85 40       STA  $0040    ;the monitor
        4C 4F1C     JMP  $1C4F    ;jump to monitor routine
        EA          NOP
        EA          NOP

        A9 C0       LDA #$C0      ;put "C0" in U10's IER to enable
        8D 0E13     STA  $130E    ;T1 interrupt

last    8A          TXA           ;get X register value
        F0 08       BEQ lin2      ;branch if 0 (2nd line input)

        A5 1A       LDA  $001A    ;else copy 1A and 1B to 15 and 16
        85 15       STA  $0015    ;to reinit. #A/D's per datum
0489    A5 1B       LDA  $001B    ;counters
        85 16       STA  $0016

lin2    98          TYA           ;get Y register value
        48          PHA           ;and save on stack
        A0 00       LDY #00       ;zero Y register.  it will be used as
                                  ;a counter here to keep track of number
                                  ;of shifts of data registers left or
                                  ;right to place appropriate values
                                  ;where they can be read

        A5 1B       LDA  $001B    ;check 1B for # A/D's per datum
        F0 21       BEQ iszero    ;to see how to roll data storage
                                  ;for reading average

        C9 80       CMP #$80      ;is it 80? (32K)
        F0 1E       BEQ is80
        C9 40       CMP #$40      ;is it 40? (16K)
        F0 1B       BEQ is40
        C9 10       CMP #$10      ;is it 10? (4K)
        F0 12       BEQ is10
        C9 20       CMP #$20      ;is it 20? (8K)
        F0 14       BEQ is20
        4C 4204     JMP  $0442    ;if none of these, fall through

        03
        4C EB05     JMP  $05EB
        18          CLC
        36 0D       ROL  $000D,X ;I think this section is obsolete!
        36 0E       ROL  $000E,X
        36 0F       ROL  $000F,X

is10    4C CB04     JMP  $04CB    ;BCD convert
iszero  C8          INY           ;store up # of times to roll data
is80    C8          INY           ;registers in Y register
is40    C8          INY
is20    C8          INY
```

```
rotat  18         CLC           ;clear carry
       76 0F      ROR  $000F,X ;rotate data registers right until
       76 0E      ROR  $000E,X ;count in Y register done
       88         DEY
       D0 F8      BNE rotat

       4C CB04    JMP  $04CB    ;BCD convert

       09
       4C D204    JMP  $04D2    ;obsolete section?!
       0A         ASL
       0A         ASL

04CB   A5 09      LDA  $0009    ;get # bytes to type and increase it
       18         CLC
       69 01      ADC #01       ;add 01
04D0   85 09      STA  $0009

04D2   68         PLA           ;restore Y register value from stack
       A8         TAY

**************************************************
this section converts HEX data to BCD.  note that 3E8 is HEX
equivalent of 1000(decimal), 64 is HEX equivalent of 100(decimal),
A is 10(decimal)
**************************************************

04D4   B5 0F      LDA  $000F,X ;check largest data register to see
       C9 03      CMP #03       ;if number may be > 1000
       90 1A      BCC nothou    ;branch if too small
       F0 12      BEQ maybe     ;branch if may be > 1000

thou   38         SEC           ;else is > 1000, set carry
       B5 0E      LDA  $000E,X ;take lower part of 1000 (Hex)
       E9 E8      SBC #$E8      ;from next lower data register
       95 0E      STA  $000E,X
       B5 0F      LDA  $000F,X ; and take high part of 1000 (Hex)
       E9 03      SBC #03       ;from highest data register
       95 0F      STA  $000F,X

04E9   E6 30      INC  $0030    ;keep track of # of 1000's in 30
       4C D404    JMP  $04D4    ;(BCD) and repeat loop

maybe  B5 0E      LDA  $000E,X ;check if next data register
       C9 E8      CMP #$E8      ;confirms data value > 1000
       B0 E8      BCS thou      ;and branch if so

nothou A5 30      LDA  $0030    ;roll up # 1000's counted in 30 to
       18         CLC           ;make way for 100's
       2A         ROL
       2A         ROL
       2A         ROL
       2A         ROL
```

```
       85 30      STA  $0030
04FD   B5 0E      LDA  $000E,X ;check this data register to see
       C9 64      CMP #$64     ;if remainder of 1000's count is > 100
       B0 06      BCS hund     ;and branch if so
       EA         NOP
       EA         NOP
       B5 0F      LDA  $000F,X ;else make sure nothing in highest
       F0 12      BEQ nohund   ;register before branching

hund   38         SEC          ;set carry
       B5 0E      LDA  $000E,X ;take 100 from 0E register and any
       E9 64      SBC #$64     ;carry from highest register
       95 0E      STA  $000E,X
       B5 0F      LDA  $000F,X
       E9 00      SBC #00
       95 0F      STA  $000F,X

       E6 30      INC  $0030   ;keep track of # of 100's in 30
0518   4C FD04    JMP  $04FD   ;and repeat loop

nohund B5 0E      LDA  $000E,X ;now take 10's out of 0E
051D   38         SEC          ;set carry
       E9 0A      SBC #$0A
       30 07      BMI noten    ;branch if no 10's
       95 0E      STA  $000E,X ;put back 0E
       E6 31      INC  $0031   ;keep track of 10's in 31
       4C 1D05    JMP  $051D   ;and repeat loop

noten  A5 31      LDA  $0031   ;roll 10's count in 31 up out of
       18         CLC          ;the way to tack on units left
       2A         ROL
       2A         ROL
       2A         ROL
       2A         ROL

0530   18         CLC          ;tack on units
       75 0E      ADC  $000E,X
       85 31      STA  $0031

       A5 30      LDA  $0030   ;move answer in 30 and 31 to data
       91 04      STA ($0004),Y;storage area indirect pointed
       C8         INY          ;to by 04+Y register
       A5 31      LDA  $0031
       91 04      STA ($0004),Y
       C8         INY

       A9 00      LDA #00      ;reinit. registers for next datum
       85 30      STA  $0030
       85 31      STA  $0031
       95 0C      STA  $000C,X
       95 0D      STA  $000D,X
0549   95 0E      STA  $000E,X
       95 0F      STA  $000F,X
```

```
        8A          TXA           ;get X register value
        84 2E       STA  $002E    ;and save Y register in 2E

        F0 30       BEQ lin2      ;branch if 0 (2nd line input)

        98          TYA           ;else get Y register
        D0 04       BNE nohi      ;branch if hi data store not to be
        E6 43       INC  $0043    ;incremented
        E6 45       INC  $0045
nohi    C6 3C       DEC  $003C    ;countdown data points
        D0 02       BNE jstlo     ;branch if not 0 (don't decrement
        C6 3D       DEC  $003D    ;high part)

jstlo   A9 40       LDA #$40      ;disable U10's T1 interrupt
0561    8D 0E13     STA  $130E

not2    AD 0D12     LDA  $120D    ;wait until end of data collection
        29 20       AND #$20      ;interval U9 T2 counts out
        F0 F9       BEQ not2

        84 42       STY  $0042    ;address of last datum in Y register
        84 44       STY  $0044    ;moved to 42 and 44

        A5 3D       LDA  $003D    ;why get 3D?
        EA          NOP
        EA          NOP

        A9 C0       LDA #$C0      ;resume A/D conversions
        8D 0E13     STA  $130E    ;enable T1 interrupt on U10

0578    A5 20       LDA  $0020    ;load U9's T2 with countdown (60Hz)
        8D 0812     STA  $1208    ;to next datum
        A5 21       LDA  $0021
        8D 0912     STA  $1209

lin2    60          RTS           ;return from subroutine
        ...........

*****************************************************
this is the start of the program when using an external trigger to
synchronize with apparatus.  there is provision for a brief delay
after trigger and before actually starting data collection.
(designed for model E UV scanner)
*****************************************************

0590    A9 00       LDA #00       ;init. accum.
        8D 0212     STA  $1202    ;U9 DDRB all inputs
        8D 0312     STA  $1203    ;U9 DDRA all ilnputs
        8D 0B12     STA  $120B    ;U9 ACR init.
        8D 0C12     STA  $120C    ;U9 PCR init.
        8D 0213     STA  $1302    ;U10 DDRB all inputs
        8D 0B13     STA  $130B    ;U10 ACR init.
        8D 0C13     STA  $130C    ;U10 PCR init.
```

```
       AD 0D13     LDA  $130D   ;check U10 IFR

05AA   A9 FF       LDA #$FF     ;all 1's
       8D 0012     STA  $1200   ;for port init's
       8D 0112     STA  $1201
       8D 0013     STA  $1300
       8D 0113     STA  $1301

       AD 0D13 not1 LDA $130D  ;check U10 IFR T1 flag
       29 40       AND #$40     ;and wait if not set
       F0 F9       BEQ not1

       AD 0413     LDA  $1304   ;U10's T1 timed out, clear its flag

not1   AD 0D12     LDA  $120D   ;do likewise for U9's T1
       29 40       AND #$40
       F0 F9       BEQ not1

       AD 0412     LDA  $1204   ;U9's T1 timed out, clear its flag

       4C D005     JMP doovr    ;look for trigger pulse train (model E)
                                ;jump can be changed to monitor if only
                                ;one scan needed
       ..............

*****************************************************
this section can be used to start the program from an externally
supplied train of trigger pulses, such as that provided by
the model E UV scanner
*****************************************************

doovr  A2 02       LDA #02      ;put "02" in X register (model E
                                ;trigger pulses to count)

       A9 E0       LDA #$E0     ;"E0" in U9's ACR for T1 interrupt and
       8D 0B12     STA  $120B   ;pulse train, T2 pulse counting
                                ;T2 will keep track of 60Hz clock

       A9 00       LDA #00      ;"00" in U10's PCR for CB1 interrupt
05D9   8D 0C13     STA  $130C   ;on hi→lo transition
                                ;CB1 is watching the triggering pulse
                                ;train from the model E

       AD 0013     LDA  $1300   ;to clear CB1 interrupt flag

nocb1  AD 0D13     LDA  $130D   ;check U10 IFR CB1 flag and wait if
       29 10       AND #$10     ;not set
05E4   F0 F9       BEQ nocb1
```

```
next    A9 01       LDA #01       ;CB1 tripped, got a trigger pulse
                                  ;from the model E
                                  ;start U9's T2 counting
        8D 0812     STA  $1208    ;60 Hz pulses
05EB    A9 00       LDA #00
        8D 0912     STA  $1209

05F0    AD 0013     LDA  $1300    ;to clear U10's CB1 flag
nocb1   AD 0D13     LDA  $130D    ;now look for U10's CB1 flag in IFR
        29 10       AND #$10      ;and wait until set
        F0 F9       BEQ nocb1     ;looking for next trigger wave

        ...EA...    NOP           ;no ops

        AD 0812     LDA  $1208    ;CB1 tripped, so check U9's T2 low
        D0 CB       BNE doovr     ;counter, branch if not timed out
                                  ;(ie false trigger from model E)
                                  ;if model E trigger wave is next in
                                  ;a series at 60Hz, the T2 should have
                                  ;had enough time to count out.  a
                                  ;fortuitous spike from the model E
                                  ;will not likely recur at 60Hz interval
                                  ;so T2 will not have counted out
                                  ;correctly.  T2 will count through zero
                                  ;to negative numbers.  resolution for
                                  ;frequency match between T2 and model
                                  ;E trigger train is 1/60 seconds

        CA          DEX           ;count the model E trigger signal
        10 DE       BPL next      ;and look for next one unless X
                                  ;register counted out

0608    4C 8B03     JMP  $038B    ;delay
```

Table AV.1

Parameters for data collection

data rate (points/min)	register values						#A/D conversions per datum
	001A	001B	001E	001F	0020	0021	
1	00	00	C7	01	0F	0E	64 K
2	00	80	C7	01	07	07	32 K
4	00	40	C7	01	83	03	16 K
5	00	40	6B	01	CF	02	16 K
10	00	20	6B	01	67	01	8 K
20	00	10	6B	01	B3	00	4 K
40	00	08	6B	01	59	00	2 K
80	00	04	6B	01	2C	00	1 K
120	00	02	E5	01	1D	00	0.5 K
180	00	02	43	01	13	00	0.5 K
240	00	01	E5	01	0E	00	0.25 K
360	00	01	43	01	09	00	0.25 K

APPENDIX VI: OPERATION OF ZIMM-CROTHERS VISCOMETER

The couette flow concentric cylinder viscometer was constructed in Eugene at the Institute of Molecular Biology, following the general design of Zimm and Crothers (1962). The inner diameter of the central glass cylinder (stator) is 7 mm with a length of approximately 60 mm. It is jacketed by a second cylinder through which temperature controlled fluid is recirculated by a Lauda K-2/R temperature bath. The cylinders are mounted on a platform over a multiratio gear motor (Geartronics) which has a magnet attached to its drive shaft and surrounding the cylinders. A thermistor probe (Yellow Springs Instruments) was inserted in the jacket for temperature measurements. Rotors were constructed from 5 mm o.d. NMR tubes, leaving a gap between rotor and stator of 1 mm to be filled by a volume of 0.8 ml of sample during measurement. An additional 1.2 ml of sample is required to load and unload the rotor from the instrument by filling the stator to a level at which the rotor can be lowered into the sample solution from above, using a triceps (3-pronged forceps, VWR Scientific). Rotor positioning and rotation are manually observed using a cathetometer (Gaertner Scientific Company). Alternatively, rotor rotation can be monitored by a laser photodiode optical assembly linked to a SUPERKIM microcomputer.

Rotors are constructed by first scoring an NMR tube striving for a break as even as possible about 3 cm from the bottom of the tube. The freshly broken end is then ground flat and perpendicular to the long axis of the tube. The tube is wrapped with Time Tape and mounted in the chuck of a drillpress with the open end extending from the chuck.

On the drillpress a wooden block is topped with a hard rubber mat and wet sandpaper which is allowed to dry as the grinding proceeds. First 240 and then 400 grit wet and dry silcon carbide paper is used followed by dry rouge paper for finishing. The open end of the rotor should be polished to a dull gloss. The rotor can be fire polished at this point, but it is my experience that correct flaming of the rotor end is fairly difficult to achieve, and unflamed rotors behave quite adequately.

Next the rotor is cleaned by rinsing it in 1N KOH in absolute ethanol. It is necessary to rinse with considerable water to completely remove the KOH. The clean, dry rotor is weighed and then placed in a beaker of solution similar in density and surface tension to that of the samples for which viscosity measurements are desired. Water is added to the rotor's interior until it floats with its open end even with the solution surface, without evidence of a meniscus. This is easily detected by backlighting the beaker and viewing the low angle reflected light from the surface of the solution around the rotor end. The rotor is removed, dried on the outside and weighed. The rotor will be loaded with an aluminum drive ring mounted concentrically with the cylinder axis, and held in place at the bottom of the inside of the rotor by RTV potting compound (available from Consolidated Electrical Supply in Portland, Oregon). The inserted drive ring and RTV8111 cement (General Electric) must add up to the weight of the water added to make the rotor float correctly. Mix the potting compound with the polymerizer and carefully add some to the drive ring area while weighing the rotor. It is my experience that the easiest way to do this is to scrape up the RTV onto a long needle

and insert the needle into the NMR tube, allowing the RTV to slowly bead up on and drip from the needle end. Repeat this procedure until the rotor is the correct weight or slightly overweight. Try to avoid getting the RTV on the inside walls since serious imbalance will cause precession of the rotor. RTV on the tube walls can be removed with a Q-tip. Let the rotor float overnight in a beaker of appropriate solution while the RTV is curing. The next morning paint two patches of Testor's black gloss enamel paint on the inside of the rotor. The paint will not stick well unless you first treat the INSIDE of the rotor with concentrated hydrofluoric acid to etch the glass. Four one minute treatments followed by extensive rinsing and drying seems adequate. The idea of the paint patches is to interrupt a laser beam to allow automatic counting of rotor rotations. The patches are also useful for counting by eye. When the paint has dried, fill the viscometer with just enough solution to float the rotor (0.8 ml in the current case). Next look at the rotor through a cathetometer and watch it at the bottom as it turns. If precession is apparent, use a long, sharp tool to carefully cut away some of the RTV. If too much is removed, more RTV must be added and allowed to cure. The goal is a rotor which does not precess.

Rotors can be characterized by the ratio of the slope to intercept of a plot of rotor speed (on the y axis) vs motor speed (on the x axis) when a Newtonian fluid is being measured. Gill and Thompson (1967) express this with the equation

$$C(\omega_m-\omega_r) + D = \eta\omega_r \qquad \text{(AVI.1)}$$

where ω_r is the angular velocity of the rotor, and ω_m is the angular velocity of the rotating magnetic field. Coefficient C is for that driving torque due to the interaction of the external magnetic field with induced eddy currents in the drive ring. The torque is a function of the rate of change of the magnetic field around the conducting drive ring, and is therefore a function of the difference in rate of rotation of the external magnet and the rotor. Term D represents driving torque due to ferromagnetic impurities in the rotor, whose moments realign as the external magnetic field changes. The resulting torque is insensitive to the rate of rotation of the magnet.

Rotors made in the above manner have characteristic ratios (C/D) of about 30. This is rather low compared to that reported for other glass rotors (Gill and Thompson, 1967). The difference is probably due to impurities in the RTV cement or aluminum stock used. The rotor is still useful for viscosity measurements as long as the effect of the impurities is corrected for. The equation for relative viscosity becomes

$$\eta_{rel} = \frac{\eta}{\eta_o} = \frac{\frac{C}{D}(\omega_m-\omega_r) + 1}{\frac{C}{D}(\omega_m-\omega_r{}^o) + 1} \times \frac{\omega_r{}^o}{\omega_r} \qquad \text{(AVI.2)}$$

where $\omega_r{}^o$ is the angular velocity of the rotor suspended in reference solution, and η_o is the reference solution viscosity.

The C/D ratio for a given rotor can actually be useful in interpretation of data, as it reflects the Newtonian character of the solution being measured. Since the non-Newtonian effect in polymer

solutions decreases with sample concentration and rate of shear, determination of C/D can establish that point at which non-Newtonian effects become negligible. For example, in measuring the viscosity of 14 μg/ml T7 DNA in solution containing 1 mM Na ion at neutral pH, the C/D in data recorded at shear rates less than 0.566 sec^{-1} is the same as that in data for buffer alone throughout the accessible shear range. The ratio thus helps establish the onset of essentially Newtonian behavior in sample solutions, as well as identifying those data which may be suspect for other reasons such as dust contamination or rotor precession.

Cleaning of the rotor exterior should be done frequently to help avoid air bubble formation and meniscus irregularities during measurements. To clean the rotor, pick it up from the side with triceps and insert it upside down in a quantity of the cleaning solution (1 M KOH in absolute ethanol) for 20-30 seconds, ensuring that the inside edge of the rotor's end also contacts the cleaning solution. Use a liberal quantity of filtered distilled water to rinse the end, and CAREFULLY remove any water inside the rotor with a fine tipped pipette or syringe with needle. Set the rotor aside to dry under the warmth of a lamp. Rotors can be supported on flat surfaces with a simple collar cut from a plastic pipette tip such as that for the Gilson Pipetteman.

It is my experience that rotors dried for a long period of time (i.e. overnight) will wet differently at the meniscus end than those dried only until no water droplets are visible. The difference appears to be in whether or not the flat polished end of the rotor wets. For consistency in measurement, rotors should always be dried

the same way. If dried for a long period of time, the meniscus will tend to form at the outside edge of the rotor end, and if for a short period, the meniscus will tend to form at the inside edge.

Before inserting the rotor into the stator, use cleaning solution on the rotor bottom which has been in contact with the collar stand, by holding the rotor with the triceps from the inside and dipping the bottom in the solution. Again use a liberal quantity of filtered water to rinse, as well as buffer and/or sample solution. Measurements should always be made with the rotor in the same position relative to the bottom of the stator, since the external magnetic field created by the rotating magnet will vary with position in the stator. This position is preferably one where rotor speed will be insensitive to small vertical displacements of the rotor, and can be established by recording rotor angular velocity as a function of its position in the stator. When the rotor has been inserted and its level is being adjusted, it is preferable to be consistent in making the final adjustment with removal of sample solution, rather than with addition. The reason for this may be regularity of meniscus formation along the wall of the stator.

APPENDIX VII: PRECOLLAPSE OF T7 DNA BY SPERMIDINE AT LOW IONIC STRENGTH: A LINEAR DICHROISM AND INTRINSIC VISCOSITY STUDY[2]

CONTRIBUTION OF COAUTHORS

The paper in this appendix presents a study performed in collaboration with Dr. W.A. Baase and Dr. S.A. Allison under the supervision of Professor J.A. Schellman at the University of Oregon. The study addresses the conformational change in DNA prior to collapse by the polyamine spermidine, by combining low-gradient linear dichroism and intrinsic viscosity data to estimate the persistence length. Dr. Allison performed the theoretical development and discussion in the paper. My contribution included the measurements of intrinsic viscosity and related sample handling. Dr. Baase isolated the T7 DNA sample material, performed linear dichroism measurements, and contributed to the discussion.

[2]Permission to include this manuscript was granted 5/16/86 by John Wiley and Sons. The manuscript was published in Biopolymers (1984), 23, p. 2835-51.

Precollapse of T7 DNA by Spermidine at Low Ionic Strength: A Linear Dichroism and Intrinsic Viscosity Study

WALTER A. BAASE, *Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403*; PAUL W. STASKUS, *Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331*; and STUART A. ALLISON,* *Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403*

Synopsis

At low salt ($[Na^+] = 10^{-3}M$), spermidine is capable of transforming DNA from a highly extended random coil to a compact particle. The transition takes place at a spermidine concentration of around 25 μM and the compact particle has been previously studied in considerable detail for several different DNAs. The objective of the present study is to see what effect, if any, spermidine has on T7 DNA conformation *prior* to collapse using flow dichroism and intrinsic viscosity. We conclude that increasing the spermidine concentration from 0 to the collapse transition point (above 20 μM) makes DNA increasingly nondraining. Furthermore, the persistence length dropped from 785 ($\pm$42) to 560 ($\pm$32) to 445 ($\pm$26) Å on increasing the ambient spermidine concentration from 0 to 1 to 10 μM. These results are in good agreement with counterion condensation theory and Odijk's theory of the electrostatic contribution to the persistence length of DNA. Nonetheless, it is concluded that counterion condensation is not entirely responsible for DNA collapse and that crosslinking promotes the transition to the compact state.

INTRODUCTION

The polyamines are all-pervasive in living systems, and the roles they play in biological systems have been dealt with extensively in the literature. The texts of Cohen[1] and Bachrach,[2] the reviews by the Tabors et al.[3,4] and the series *Advances in Polyamine Research* can be cited as works on the subject of general interest. In particular, the polyamines have a substantial effect on DNA flexibility and conformation that is important in connection with the fact that most DNA in living systems is present in a compact form. In eukaryotic cells, DNA is compacted into chromatin by histones and, in virions, DNA exists as a compact, possibly toroidal[5] mass surrounded by a protein

* Present address: Department of Chemistry, Georgia State University, Atlanta, Georgia 30303.

Biopolymers, Vol. 23, 2835–2851 (1984)

CCC 0006-3525/84/122835-17$04.00

shell or capsid. In the last 12 years, a number of chemical environments have been discovered that mimic this compaction *in vitro*. Polyethylene oxide,[6,7] polylysine,[8,9] polyethylene glycol,[10] a variety of aqueous salt–alcohol systems,[11-14] cobalt hexamine,[15,16] and polyamines[11,14,17-26] are capable of compacting DNA and have been studied by a variety of techniques, including electron microscopy, flow linear dichroism (LD), and static, dynamic, and electrophoretic light scattering.

Despite a considerable amount of research on the compact form of DNA, little is known regarding its conformation prior to collapse. For example, in an ambient, aqueous salt solution containing $[Na^+] = 10^{-3}M$, the polyamine spermidine collapses DNA into a compact structure above a concentration of around 25 μM and the compact structure has been characterized.[14,17-23,25] Nonetheless, DNA conformation prior to collapse, between $0 < [sp] \leqslant 20\ \mu M$, has not been studied in any detail, although light-scattering results indicate a random-coil configuration.[14,22] At a somewhat higher salt concentration of $[Na^+] = 10^{-2}M$, where DNA precipitation, rather than collapse, occurs at a spermidine concentration in excess of 150 μM, dynamic light scattering has shown that viral ϕ29 DNA remains in the random-coil configuration up to the precipitation point.[26] Furthermore, the radius of gyration has been found to decrease from 280 to 190 nm on raising the spermidine concentration from 0 to 40 μM. It can be expected that similar, though more dramatic, changes take place at $[Na^+] = 10^{-3}M$, and this is the subject of the present work.

In this work, the combined techniques of flow LD and Zimm-Crothers viscometry are used to study T7 DNA in $[Na^+] = 10^{-3}M$ with spermidine present in the concentration range of 0–20 μM. The experiments and the interpretation of the results are based on a methodology developed in Schellman's laboratory over the past seven years.[27,28] This has involved the development of a flow instrument that makes use of modern polarization–modulation techniques, the solution of the problem of the LD of a wormlike chain, and the incorporation of these results into hydrodynamic theory.[27-29] Measurements of the low-gradient viscosity and the LD of a stream-oriented flexible molecule provide a measure of the flexibility of the molecule. As described later, LD measured 45° from the direction of flow at very low gradients gives information regarding equilibrium properties of DNA unperturbed by the flow field, while LD measured at zero degrees from the direction of flow indicates the extent to which the molecule is stretched by the flow field. Perhaps the most important equation in this work is Eq. (5), which relates the low-gradient LD to the intrinsic viscosity of DNA, $[\eta]$, the ratio of the persistence length, P_∞, to the contour length, L, and the reduced LD of a base pair, $(\Delta\epsilon/\epsilon)_{bp}$. If the LD and intrinsic viscosity of a monodisperse sample are measured, then the product, $P_\infty(\Delta\epsilon/\epsilon)_{bp}$, is uniquely determined.

MATERIALS

Spermidine-trichloride was purchased from Sigma and purified by double recrystallization from ethanol. All measurements were made using bacteriophage T7 DNA grown and purified according to standard procedures,[30] ending with hot phenol extraction ($T = 55$°C). The stock DNA was exhaustively dialyzed against 0.1M Tris buffer containing $10^{-3}M$ EDTA (pH 7.4) and stored in the refrigerator (4°C) at an optical density of between two and five. Over the total length of time that elapsed during the course of these experiments (about 1½ years), four different DNA stocks were prepared and used in the LD and viscosity experiments. Boundary sedimentation velocity determinations were routinely made using a Beckman model E analytical ultracentrifuge equipped with scanner and absorption optics. These sedimentation velocity determinations were carried out at 0.2M NaCl, $10^{-4}M$ EDTA, $0.9 \times 10^{-3}M$ sodium phosphate (pH 7.2) and extrapolated to the expected value of 31.6 S at infinite dilution.[31,32] Single sharp boundaries were observed, indicative of a monodisperse sample. When a new DNA stock was prepared, it was checked for reproducibility by repeating an earlier LD or viscosity experiment under identical conditions.

The low-salt buffer used in all the experiments consisted of 0.1 mM trisodium EDTA plus 0.35 mM Na_2HPO_4 (pH $\simeq$ 7.4) in filtered, double-distilled water to give an ambient sodium ion concentration of $10^{-3}M$. Stock DNA aliquots were diluted to between 0.8 OD/mL (for viscosity) or 0.08 OD/mL (for LD) with low-salt buffer and repeatedly dialyzed against at least 40-fold greater volumes of low-salt buffer containing spermidine. At the lower spermidine concentrations, additional spermidine was added to the *first* dialysis solution, since approximately 80% of the DNA phosphate groups are "condensed" by spermidine on the basis of counterion condensation theory.[33-35] Repeated changes of the dialysis bath eliminated this complication. Dialysis was carried out for 36–48 h at 4°C.

Prior to an experiment, the sample was diluted to 0.04 OD/mL for LD and 0.3 OD/mL for viscosity with dialysate and was allowed to sit at room temperature for 1–2 h, followed by brief vacuum degassing to prevent air-bubble formation. Both the LD and viscosity experiments were carried out at 25 $\pm$ 0.02°C. DNA concentration was determined using a Cary 15 spectrophotometer with 10-cm-pathlength cells for LD measurements and a Cary 219 spectrophotometer with 1-cm-pathlength cells for viscosity measurements. A molar extinction coefficient of 6600 at 257 nm was assumed for T7 DNA.

In the case of the LD measurements, fresh samples were prepared for each spermidine concentration studied and discarded at the end of the experiment. This precaution was taken to minimize contamination of the sample by atmospheric dust or contact with the stainless-

steel stator. The LD measurements were reproducible throughout a particular experiment (which usually lasted about 3 h). When not in use, the LD sample cell was washed by continuous flow with double-distilled water. Measurements of the conductivity of double-distilled water before and after being placed in the clear sample cell were identical.

METHODS

Linear Dichroism

The LD apparatus is described in detail elsewhere.[27,28,36] The DNA solutions were oriented in a Couette flow cell consisting of two concentric cylinders about 11 cm in length separated by a 0.1-cm gap in which the sample is placed. The outer cylinder is rotated with a variable speed motor that produces a linear velocity gradient across the sample. Measurements were made with light propagating through the gap parallel to the axis of rotation of the cylinders with a pathlength of 10 cm. The light source is a low-pressure mercury lamp combined with an interference filter to isolate the emission at 253.7 nm. The light beam is modulated by a stress modulator-Rochon polarizer combination that produces light that switches between two orthogonal polarizations at 100 kHz.[37] The detected alternating component of the transmitted intensity is directly proportional to the LD of the sample, ΔA, where

$$\Delta A_\theta = (A_\theta - A_{\theta \pm 90^\circ}) \tag{1}$$

and θ is measured relative to the direction of flow of the solution. θ may be varied by rotating the modulator–polarizer unit about an axis parallel to the light beam. The proportionality constant is determined by calibrating with a Rochon prism.

Traditionally, the two independent quantities measured in LD are the angle of isocline, χ, which is defined as the angle between the principal axis of the refractive index ellipsoid and the direction of flow and ΔA_χ, the LD measured at the angle χ. Instead, our practice is to measure ΔA_0 and ΔA_{45}, where 0 and 45 refer to the angular setting of the polarizer–modulator relative to the flow direction. These can be related to χ and ΔA_χ by

$$\cot(2\chi) = \Delta A_0 / \Delta A_{45} \tag{2}$$

$$\Delta A_\chi = (\Delta A_0^2 + \Delta A_{45}^2)^{1/2} \tag{3}$$

It turns out that ΔA_0 and ΔA_{45} are easier and faster to measure, and

are also more simply related to theory for both flexible[38] and rigid[39] macromolecules.

Flow data was processed essentially according to Rizzo and Schellman,[28] except that a PDP 11/23 computer system and a new FORTRAN program are now used to collect the flow data and electronically count rotations of the flow cell. The values of $\Delta A_{45}/A$ were fitted to the quadratic polynomial $a_{45}G + b_{45}G^2$ and $\Delta A_0/A$ to the cubic polynomial $a_0G^2 + b_0G^3$, where G is the velocity gradient in s^{-1}. The limiting slopes can be identified with a_{45} and a_0, respectively, which are required in the data analysis described later. Further details regarding the procedure can be found elsewhere.[27,28]

Intrinsic Viscosity

Relative viscosities were measured at 25°C in a semimicro version of the Zimm-Crothers floating rotor viscometer.[40] The inner diameter of the stator was 7 mm, and the rotor was made from 5-mm outer diameter nmr tubes with an aluminum drive ring. Measured relative viscosities were corrected for the contribution of ferromagnetic impurities in the rotor according to Gill and Thompson.[41] An optical system was used to count rotor rotations. It employed a 0.5-mW demonstration laser, photosensitive transistor connected to a 6-V battery and a single-board 6502-based microcomputer (Lamar Instruments, Redondo Beach, CA), with an analog-to-digital converter. Laser light aimed through the viscometer temperature bath, stator, and rotor was periodically intercepted by enamel patches painted inside the rotor. The transmitted light was detected by the transistor, which then passed voltage to the A/D converter. The computer was programmed in machine code to monitor the changing voltage level and to time intervals with a 60-Hz clock.

Measurements were made at four different gradients with shear stresses from 1.92×10^{-3} to 14.6×10^{-3} dyn/cm^2. Rotor and stator were rinsed thoroughly with the appropriate dialysis buffer before the DNA samples were placed in the stator. The DNA samples were introduced using 0.38-mm-i.d. polyethylene tubing connected to a syringe. After measurements at four different gradients were carried out, the sample was removed and its concentration determined optically using a Cary 219 spectrophotometer. In all cases, three to five different DNA concentrations were investigated. The intrinsic viscosity was then determined using the well-known relation[42]:

$$[\eta] = \lim_{c \to 0} (\eta_s - \eta_0)/\eta_0 c \tag{4}$$

where c is the DNA mass concentration, η_0 the buffer viscosity, and η_s the sample viscosity. The data were also plotted as $\ln(\eta_{rel})/c$ vs c, where η_{rel} is the relative viscosity.

INTERPRETATION OF THE LD MEASUREMENTS

In a previous publication,[27] the Zimm theory for birefringence of flexible macromolecules[38] has been transformed into the appropriate expressions for LD. Chain averages have also been performed, with the result that the dichroic properties of the monomer unit appear explicitly in the expressions for the LD. In the case of a long, stiff macromolecule such as linear double-stranded DNA, the low-gradient limits of ΔA_{45} and ΔA_0 are given by[28]

$$\frac{\Delta A_{45}}{A} = \frac{2}{3}\frac{\Delta\epsilon_{bp}}{\epsilon_{bp}}\frac{P_\infty}{L}\beta \tag{5}$$

$$\frac{\Delta A_0}{A} = \frac{2}{3}\frac{\Delta\epsilon_{bp}}{\epsilon_{bp}}\frac{P_\infty}{L}\beta^2 K \tag{6}$$

where A is the isotropic orientable absorbance, P_∞ the persistence length, L the contour length of the macromolecule, ϵ_{bp} the isotropic extinction coefficient of a base pair [$\epsilon_{bp} = \epsilon_\parallel + 2\epsilon_\perp$, where $\epsilon_\parallel$ and $\epsilon_\perp$ are the monomer extinction coefficients parallel and perpendicular to the propagation (helix) axis respectively], and $\Delta\epsilon_{bp} = \epsilon_\parallel - \epsilon_\perp$. The quantity β is the Peterlin gradient parameter,

$$\beta = (\eta_0[\eta]M/RT)G \tag{7}$$

where G is the gradient, η_0 the solvent viscosity, $[\eta]$ the intrinsic viscosity, M the molecular weight of the macromolecule, R the gas constant, and T the absolute temperature. Finally, the quantity K has been identified by Tschoegl[43] as the reduced steady-state compliance of the macromolecule. Roughly speaking, K measures the strength of the intramolecular hydrodynamic interaction. $K = 0.4$ for a free-draining Rouse-Zimm chain and $K \simeq 0.20$ for a completely nondraining chain.

It is worthwhile at this point to comment on the nature and range of validity of Eqs. (5) and (6). It turns out that Eq. (5) pertains to the equilibrium conformation of the macromolecule unperturbed by the hydrodynamic field. Physically, a nonvanishing ΔA_{45} results from the tendency of the end-to-end vector (or longest Rouse-Zimm mode) to orient at 45° relative to the flow direction rather than from an actual deformation of the macromolecule. Consequently, equilibrium chain statistics can be used to relate ΔA_{45} to the model parameters. ΔA_0, on the other hand, gives a direct measure of the extensibility of the macromolecule. Equation (6) can also be written[29,44]

$$\frac{\Delta A_0}{A} = \frac{\Delta\epsilon_{bp}}{\epsilon_{bp}}\frac{P_\infty}{L}\left(\frac{\langle r^2\rangle - \langle r^2\rangle_0}{\langle r^2\rangle_0}\right) \tag{8}$$

where $\langle r^2 \rangle_0$, $\langle r^2 \rangle$ is the rms end-to-end distance in the absence or presence of a gradient, respectively.

Although Eqs. (5) and (6) were derived using the bead-spring model of Zimm, the works of Gottlib and Svetlov[45] and Kramers[46] have shown that these equations have wide applicability and are not subject to many of the restrictions of the bead-spring model. For example, Eq. (5) applies to the wormlike coil model with hydrodynamic interaction but no excluded volume.[47,48] For high-molecular-weight DNA, excluded-volume effects are expected to be significant. We shall assume that the treatment of excluded volume of a bead-spring polymer using the ϵ parameter proposed by Ptitsyn and Eizner[49] and developed by Tschoegl[43,50,51] and Bloomfield and Zimm[52] can be applied in our case. The theory shows that Eqs. (5) and (6) are unchanged when excluded-volume effects are present. Equation (5) depends implicitly on the chain-expansion parameter, ϵ. In other words, ΔA_{45} and $[\eta]$ have the same dependence on ϵ and remain proportional to one another. A review of the theories shows that the expansion parameter is introduced so that it affects *inter*bead dimensions, and that the Zimm length parameter, b^2, is not considered to be a function of ϵ. In using these theories then, long-range effects of spermidine are considered to be described by ϵ and manifest themselves in a change in $[\eta]$, whereas short-range effects give rise to a change in persistence length. A more in-depth analysis of this problem has been given previously.[28]

Equation (6) has a more restricted range of validity than Eq. (5) and depends on the Ptitsyn-Eizner-Tschoegl-Bloomfield-Zimm[49–52] analysis of the bead-spring model with excluded-volume and hydrodynamic interaction. Since it involves a stretching of the molecule from equilibrium dimensions, an adequate description requires a dynamic theory of perturbation by flow. The Zimm theory provides such a description.

RESULTS

In the viscosity experiments, intrinsic viscosities were determined in $[Na^+] = 10^{-3}M$ buffer plus the following concentrations of ambient spermidine: 0, 1, 4, and 10 μM. The corresponding intrinsic viscosities were determined to be 357, 250, 213, and 205 dL/g, respectively, with an estimated error of $\pm 5\%$. Intrinsic viscosities for other spermidine concentrations were calculated by interpolation or extrapolation. A plot of $\log[\eta]$ versus the logarithm of the free spermidine concentration was found to be linear. Specifically, $\log[\eta] = -0.09017 \times \log[\text{sp}] + 1.8526$, where $[\eta]$ is in dL/g and [sp] is molar. Our experience has been that the greatest source of random error was the determination of the DNA concentration at high dilution. A Cary 219 spectrophotometer had sufficient sensitivity and long-term stability to provide good concentration data. This was, however, not the case for the same samples

in a Cary 14 or 15. Occasionally, rotor precision and/or air bubbles on the rotor produced fairly large systematic errors and such data were discarded.

Representative results were plotted in Fig. 1 as η_{sp}/c vs c and as $\ln(\eta_{rel})/c$ vs c. In our instrument, viscosity data were linear with respect to T7 DNA concentration below 15 μg/mL. Furthermore, below 15 μg/mL, extrapolation to zero concentration gave the same intercept to within 1% for both types of plots.

Typical LD results are shown in Figs. 2 and 3. The measured values of the reduced dichroism $\Delta A_{45}/A$ and $\Delta A_0/A$ versus the velocity gradient are plotted from the lowest (▲ – [sp] = 0 μM) to the highest (■ – [sp] = 20 μM) concentration of spermidine present in all samples studied. In both figures, the center data is for intermediate concentration ([sp] = 1 μM). In all of the LD experiments, the DNA con-

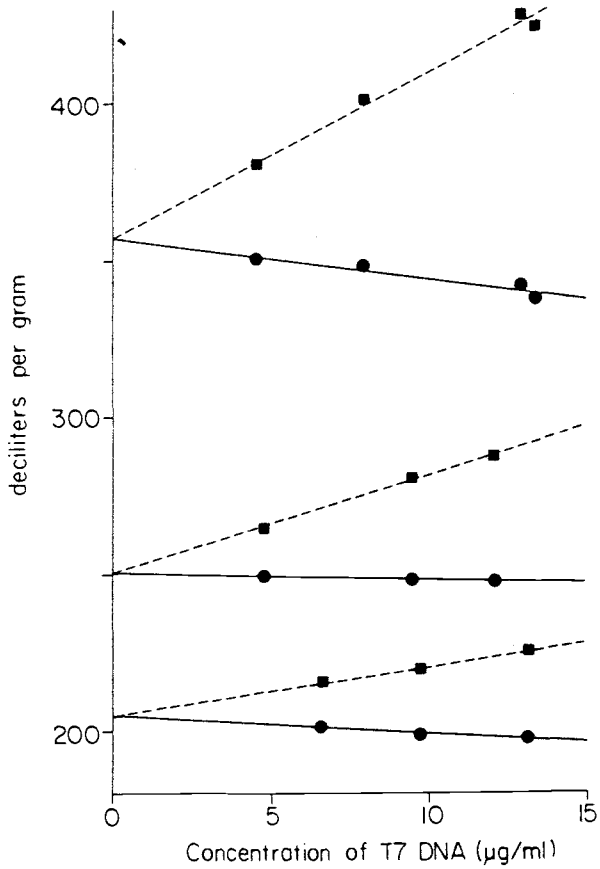

Fig. 1. Viscosity concentration dependence of T7 DNA in 0.35 mM Na_2HPO_4–0.1 mM Na_3EDTA at several spermidine concentrations. The filled squares (■) are data plotted as η_{sp}/c vs c. The filled circles (●) are the same data plotted as $\ln(\eta_{rel})/c$ vs c. Data sets are 0, 1, and 10 μM spermidine from the top to bottom.

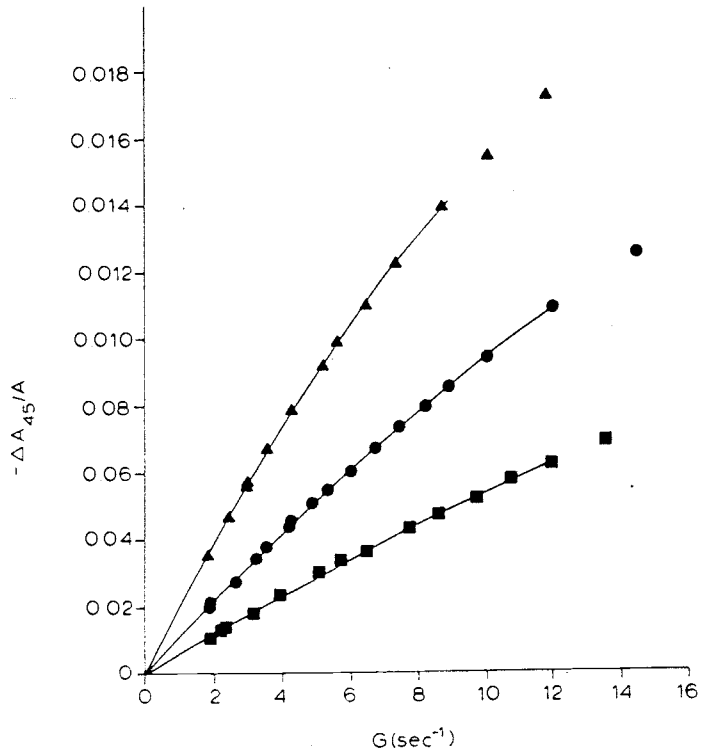

Fig. 2. Velocity-gradient dependence of $\Delta A_{45}/A$ for T7 DNA at various spermidine concentrations. The lines are the least-squares fit to $y = ax + bx^2$. The spermidine concentrations are 0.0 (▲), 1.0 (●), and 20.0 (■) μM.

centration was extremely low, lying between 2 and 2.5 μg/mL. In general, the errors in $\Delta A_{45}/A$ and $\Delta A_0/A$ were comparable to the size of the symbols in the figures. As in a previous study,[28] the gradient regime of greatest interest is from about 3 to 10 s^{-1}. At gradients below 3 s^{-1}, weak signals and fluctuations in the motor speed produce unreliable results, and at gradients above about 10 s^{-1}, the low gradient regime for which Eqs. (5) and (6) are valid breaks down.

Table I summarizes the results of this work. $\Delta A_{45}/AG$, $\Delta A_0/AG^2$, and $d\cot(2\chi)/dG$ represent low-gradient limits, and the quantities in parentheses are the estimated standard errors. The quantity K was determined using the relation

$$K = \frac{1}{\beta'}\frac{d\cot(2\chi)}{dG} \tag{9}$$

where $\beta' = \beta/G = \eta_0[\eta]M/RT$. It should be noted that the angle of isocline and also K are independent of the optical base-pair anisotropy, $\Delta\epsilon_{bp}/\epsilon_{bp}$. The persistence length, on the other hand, was determined

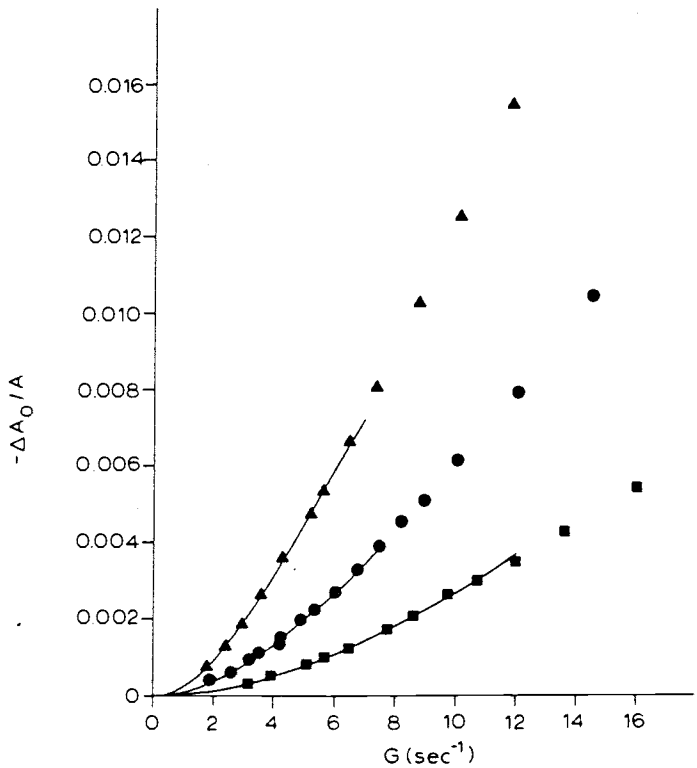

Fig. 3. Velocity-gradient dependence of $\Delta A_0/A$ for T7 DNA at various spermidine concentrations. The lines are the least-squares fit to $y = ax^2 + bx^3$. The spermidine concentrations are 0.0 (▲), 1.0 (●), and 20.0 (■) μM.

from Eq. (5) using the assumed value of $\Delta\epsilon_{bp}/\epsilon_{bp} = -1.4$ for B-form DNA. Using 39,936 base pairs as the degree of polymerization[32] and a monomer length of 0.338 nm, we obtain a contour length, L, of 1.35×10^4 nm for T7 DNA.

DISCUSSION

Figure 4 shows how K varies with spermidine concentration. Although the value of $K = 0.335$ for the [sp] = 0 sample is higher than expected on the basis of earlier work done in this laboratory[28] ($K = 0.263$ for a similar T7 sample but with $[Na^+] = 0.002M$ rather than $0.001M$), the higher K value obtained in the present work agrees qualitatively with the findings of Harrington in his study of flow birefringence of T2 DNA ($M_w = 130 \times 10^6$).[53] He reports a change of K from 0.222 at 1–2M $[Na^+]$ up to 0.377 at 0.005M $[Na^+]$. Increasing the spermidine concentration has the effect of reducing K, which can be interpreted as inducing DNA to adopt an increasingly nondraining state. This is one of the main conclusions of the paper and is expected,

TABLE I

Opticohydrodynamic Properties of T7 DNA at $10^{-3}M$ Na^+ Ion as a Function of Ambient Spermidine Concentration

[sp] (μM)	$-\frac{\Delta A_{45}}{AG}$ ($\times 10^3$)[a]	$-\frac{\Delta A_0}{AG^2}$ ($\times 10^4$)[a]	$\frac{d \cot 2\chi}{dG}$ ($\times 10$)[a]	[η] (dL/g)	β'[b]	K	P_* (Å)
0	2.056(0.02)	2.605(0.02)	1.267(0.016)	357(15)	0.379(0.02)[c]	0.335(0.018)	785(42)
1	1.112(0.02)	1.044(0.04)	0.939(0.04)	250(10)	0.287(0.015)[d]	0.327(0.022)	560(32)
4	0.885(0.015)	0.674(0.04)	0.762(0.05)	213(9)	0.244(0.012)[d]	0.312(0.025)	524(28)
7	0.806(0.01)	0.573(0.03)	0.711(0.04)	208(8)[e]	0.238(0.012)[d]	0.299(0.022)	490(26)
10	0.723(0.02)	0.474(0.04)	0.655(0.06)	205(8)	0.235(0.012)[d]	0.278(0.03)	445(26)
20	0.606(0.01)	0.349(0.02)	0.576(0.03)	189(8)[f]	0.217(0.01)[d]	0.265(0.02)	404(22)

Note that the numbers in parenthesis refer to absolute, not relative, error limits.

[a] In the limit $G \to 0$.

[b] $\beta' = \eta[\eta_0]M_r/RT$.

[c] $M_r = 657 \times 39936$ for Na-DNA.

[d] $M_r = 710 \times 39936$ for 90% spermidine DNA in sodium buffer; $I = 298.2$ K.

[e] Interpolated from results at 1, 4, and 10 μM spermidine; $\log[\eta] = -0.09017 \times \log_{10}[\text{sp}] + 1.8526$.

[f] Extrapolated from results at 1, 4, and 10 μM spermidine.

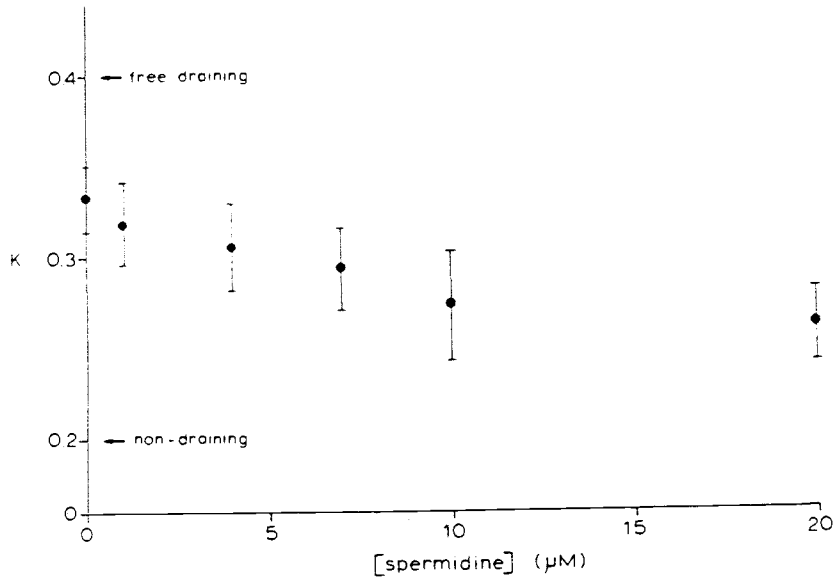

Fig. 4. Dependence of K on spermidine. Arrows denote free and nondraining limits for a bead-spring mode. Data from Table I.

since spermidine ultimately causes DNA to collapse. In the collapsed state, DNA is expected to be nondraining.

The methodology used in this work actually determines the *product* of the base-pair anisotropy and the persistence length and does not allow their separate evaluation. This is a general feature of flow methods,[27,28,53,54] and complicates *quantitative* persistence length determinations, since dichroism and birefringence measurements are very sensitive to optical anisotropies that are not known to high precision.[27,28,54] $\Delta\epsilon_{bp}/\epsilon_{bp}$ is approximately equal to -1.4, which is the value used in this work. This number is based on extensive theoretical calculations using the Arnott structure of B-DNA,[55] which shall be reported elsewhere. In any case, the persistence lengths given in Table I would have to be scaled by a multiplicative constant in the event the base-pair anisotropy turns out to be different from -1.4. This, however, would not be the case in the event of a spermidine-induced change in the base-pair anisotropy. Fortunately, there is strong, but indirect, evidence that such a spermidine-induced optical change does not occur. Using a parallel plate, high-shear flow apparatus, Johnson, Jr. and coworkers have found that simple monovalent salts do not induce any noticeable change in $\Delta\epsilon_{bp}/\epsilon_{bp}$ of DNA up to salt concentration of $1M$.[56] On the basis of CD experiments, the addition of polyamines is qualitatively similar to increasing the ionic strength, al-

though the effect is seen at much smaller polyamine concentrations.[57] Hence, it is reasonably safe to assume that there is no spermidine-induced change in the base-pair anisotropy.

Figure 5 shows how the measured persistence length varies with spermidine concentration (filled circles). The solid and dotted lines in Fig. 4 are theoretical curves based on the model of Odijk[58-60] and others.[61,62] In this model, the persistence length can be written as the sum of two parts

$$P_x = P_x^0 + P_x(\text{elect}) \tag{10}$$

where P_x^0 is the "intrinsic" persistence length and is due to the finite rigidity of uncharged DNA, and $P_x(\text{elect})$ represents the contribution of the negative phosphate charges on the DNA backbone. Since these charges repel each other, $P_x(\text{elect})$ is positive. At a Na^+ concentration of $10^{-3}M$, Eq. (10) can be written[60]

$$P_x = P_x^0 + 5700(1 - r)^2 \tag{11}$$

where the units are angstroms and r is the fraction of DNA phosphates that have associated or "condensed" cations. The dotted line in Fig. 5 was obtained using Eq. (11) with r calculated using counterion condensation theory as modified by Wilson and Bloomfield for two simple

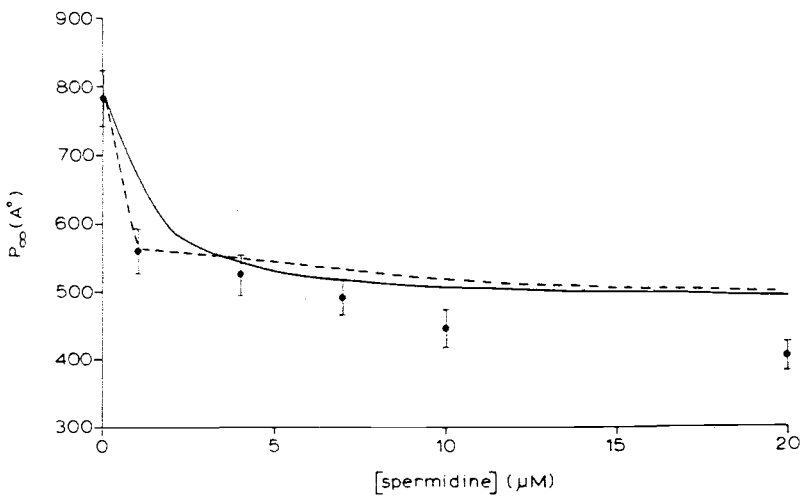

Fig. 5. Dependence of P_x on spermidine concentration. Filled circles with error bars denote experimental points and the solid and dashed lines are theoretical predictions based on experimental binding studies and counterion condensation theory, respectively. Data from Table I.

cations.[21] The solid line was obtained using Eq. (11) with r calculated using the results of experimental binding studies carried out by Braunlin et al.[63] In the absence of spermidine, we chose $r = 0.76$, as predicted by condensation theory. In addition, a value of $P_\infty^0 = 470$ Å was employed. This value was obtained in an earlier study[28] using the same methodology and DNA (T7) as used in the present work. As can be seen, the measured and calculated persistence lengths are in good agreement with each other, provided the spermidine concentration is less than about 10 μM. The solid curve calculated from the experimental binding study is an upper bound, particularly at low spermidine concentrations, since the contribution of Na^+ to r has been neglected (except at [sp] = 0).

Up to a concentration of around 10 μM spermidine, the behavior of the persistence length can be explained rather well by counterion condensation theory.[33-35] Spermidine, being a trication, condenses on the DNA to a greater extent than monovalent Na^+, and r increases suddenly from about 76% (no spermidine) to about 87% when [sp] = 1 μM. This results in a substantial decrease in electrostatic repulsion, which accounts for the large drop in P_∞ between 0 and 1 μM spermidine. A larger concentration of free spermidine results in a slightly larger r value and, consequently, a lower P_∞, but the largest change takes place at low spermidine concentrations.

The [sp] = 20 μM result does not appear to agree nearly as well with the theory as do the results up to about [sp] = 10 μM, and the probable reason for this is that the DNA is no longer adequately described in terms of a random-coil or Gaussian-chain model. In an earlier study under almost identical solvent conditions, light scattering from viral ϕ29 DNA at [sp] = 20 μM was best interpreted as resulting from an inhomogeneous suspension of partially condensed DNAs.[22] At a concentration of [sp] = 30 μM, the transition to a collapsed structure is complete.

The picture that emerges, then, is that counterion condensation *alone* is not entirely responsible for the collapse of DNA. Counterion condensation reduces substantially the electrostatic repulsion between the negative charges on the DNA backbone, but this condensation occurs primarily in the range 0 < [sp] < 10 μM, *where DNA remains in the random-coil configuration.* Actual collapse does *not* occur for this system unless [sp] ≥ 25 μM.[17,18,21-23] Evidence is growing that DNA collapse in low-salt polycation systems depends on the specific counterion as well as its charge.[21-24] In addition to counterion condensation, the formation of counterion-induced crosslinks at high counterion binding levels is probably needed to promote the transition to a collapsed state.[22-24] Recent x-ray diffraction studies of DNA–counterion fibers[64] support the crosslinking hypotheses. When, for example, the butyl moiety of spermidine is replaced by propyl through octyl moieties, the interhelical strand separation varies in a systematic way that is readily explained in terms of a crosslinking model.

This conclusion is not in agreement with an early hypothesis of Manning[35] that the random-coil configuration of DNA is intrinsically unstable ($P^0_\infty < 0$) and that this configuration is maintained by electrostatic repulsions. Although the empirical rule of Wilson and Bloomfield[21] that DNA collapse occurs when 88–90% of the DNA phosphates are neutralized by bound counterions lends support to Manning's simple and appealing hypothesis, Widom and Baldwin[16] cite several cases that do not obey this empirical rule. In the spermidine–DNA system studied in this work, 90% of the DNA phosphates are neutralized by bound counterions when [sp] = 5.8 μM, according to the experimental binding studies of Braunlin et al.[63] This is well below the concentration required to cause collapse. The observation that the rule of Wilson and Bloomfield is a good, but approximate criterion for predicting whether DNA collapses under widely varying solvent conditions is undoubtedly a consequence of specific ion effects modifying the dominant electrostatic interactions between DNA and associated counterions.

Recently, Manning has reconsidered the problem and now concludes that $P^0_\infty = 317$ Å.[65] Nonetheless, we feel this result for P^0_∞ is still too low and that his theory predicts too large a counterion concentration effect on the persistence length.

An important question that remains unresolved is the separation of long-range excluded volume, manifested in the chain-expansion parameter, ϵ, from short-range stiffness manifested in P_∞. This aspect of the analysis has been discussed in an earlier section of this work and in Rizzo and Schellman's study of T7 DNA as a function of sodium ion concentration.[28] In their analysis of the reduced steady-state compliance, K, for example, the draining parameter, h, was found to decrease with increasing salt concentration, which is opposite the expected trend. Since K is extracted from the ΔA_0 data, its rigorous interpretation in terms of the bead-spring model depends on the adequacy of both the static and dynamic bead-spring theories, as well as the ϵ-parameter theories of excluded volume. Because of this, it was unclear whether the ΔA_0 data reflected the inadequacy of the dynamic bead-spring model or the ϵ-parameter theories. Their analysis of $[\eta]$ and ΔA_{45}, which does not depend on the dynamic bead-spring model, yielded reasonable results if the DNA was allowed to be partially free-draining.

A parallel analysis of the present data has been carried out, but the results are less conclusive than those of Rizzo and Schellman. The reason for this is that in a strict interpretation of the data in terms of the bead-string model, two parameters, ϵ and h (the draining parameter), have to be either specified or determined. Rizzo and Schellman were able to specify ϵ, using an empirical equation derived by Douthart and Bloomfield,[66] $\epsilon = 0.05$–$0.063 \log[Na^+]$, which left only the draining parameter to be determined. In this work, on the other hand, both h and ϵ are unknown except for the [sp] = 0 case.

SUMMARY

The techniques of flow LD and intrinsic viscosity have been combined to study the effect of spermidine on the random-coil configuration of DNA. It is concluded that spermidine makes high-molecular-weight DNA increasingly nondraining and reduces the persistence length. The reduction in persistence length is well explained using the persistence-length model of Odijk and others,[58-62] along with counterion condensation theory[21] or experimental binding studies[63] to calculate the fraction of "condensed" DNA phosphates. At the same time, it is concluded that counterion condensation alone is not responsible for DNA collapse. Crosslinking is cited as a probable promoter to the collapsed state.[22,24,64]

Professor John A. Schellman deserves special thanks for his supervision and support of all phases of this work. In addition, Professor W. Curtis Johnson, Jr. provided support at Oregon State University (OSU), and the following people deserve acknowledgement for assistance in the preparative work: Professor Timothy M. Lohman, Jim R. Wirkkula, Jr., and Elizabeth Pluhar. This research was funded in part by Grants GM 20195 from the U.S. Public Health Service (J.A.S), PCM 8104339 from the National Science Foundation (J.A.S.), and, in part, by Grant PCM 8021210 from the National Science Foundation (to W.C.J.) and by Fellowship No. 1 F32 GM08090-01 from the National Institutes of Health (to S.A.A.). We would also like to thank Professor Michael Schimerlick (OSU) for the use of his model 219 Cary spectrophotometer, which was of critical importance during our viscosity measurements, and to thank Ms. Mary Gilland for her secretarial and editorial work.

References

1. Cohen, S. (1971) *Introduction to the Polyamines*, Prentice-Hall Inc., Englewood Cliffs, N.J.
2. Bachrach, U. (1973) *Function of Naturally Occurring Polyamines*, Academic Press, New York.
3. Tabor, H., Tabor, C. W. & Rosenthal, S. (1961) *Annu. Rev. Biochem.* **30**, 579–604.
4. Tabor, C. W. & Tabor, H. (1976) *Annu. Rev. Biochem.* **45**, 285–306.
5. Richards, K. E., Williams, R. C. & Calendar, J (1973) *J. Mol. Biol.* **78**, 255–259.
6. Lerman, L. S. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 1889–1890.
7. Lerman, L. S. (1973) *Cold Spring Harbor Symp. Quant. Biol.* **38**, 59–73.
8. Shapiro, J. T., Leng, M. & Felselfeld, G. (1969) *Biochemistry* **8**, 3219–3232.
9. Laemmli, U. K. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 4288–4292.
10. Pyatigorskay, T. L. Evidokimov, Y. M. & Varshavski, Y. M. (1978) *Mol Biol. (Moscow)* **12**, 404–412.
11. Eichbush, T. H. & Mondrianakis, E. N. (1976) *Cell* **13**, 295–306.
12. Lang, D. (1973) *J. Mol. Biol.* **78**, 247–254.
13. Lang, D., Taylor, T. N., Dobyan, D. C. & Gray, D. M. (1976) *J. Mol. Biol.* **106**, 97–107.
14. Post, C. B. & Zimm, B. H. (1982) *Biopolymers* **21**, 3123–3160.
15. Widom, J. & Baldwin, R. W. (1980) *J. Mol. Biol.* **144**, 431–453.
16. Widom, J. & Baldwin, R. L. (1983) *Biopolymers* **22**, 1595–1620.
17. Gosule, L. C. & Schellman, J. A. (1976) *Nature* **259**, 333–335.
18. Chattoraj, D. K., Gosule, L. C. & Schellman, J. A. (1978) *J. Mol. Biol.* **121**, 327–337.
19. Campbell, R. A. et al., Eds. (1979) *Advances in Polyamine Research*, Vol. 1, Raven Press, New York.
20. Skuiridin, S. G., Kadykov, V. A., Shashkov, V. S., Evdokimov, Y. M. & Varshavski, Y. M. (1978) *Mol. Biol. (Moscow)* **12**, 413–420.

21. Wilson, R. W. & Bloomfield, V. A. (1979) *Biochemistry* 18, 2192–2196.
22. Allison, S. A., Herr, J. C. & Schurr, J. M. (1981) *Biopolymers* 20, 469–488.
23. Thomas, T. J. & Bloomfield, V. A. (1983) *Biopolymers* 22, 1097–1106.
24. Yen, W. S., Rhee, K. W. & Ware, B. R. (1983) *J. Phys. Chem.* 87, 2148–2152.
25. Ruben, G. C., Marx, K. A. & Reynolds, T. C. (1981) *39th Annual Proceedings of the Electron Microscopy Society of America*, Bailey, G. W., Ed., Atlanta, pp. 438–441.
26. Lin, S. C., Thomas, J. C., Allison, S. A. & Schurr, J. A. (1981) *Biopolymers* 20, 209–230.
27. Wilson, R. W. & Schellman, J. A. (1978) *Biopolymers* 17, 1235–1248.
28. Rizzo, V. & Schellman, J. A. (1981) *Biopolymers* 20, 2143–2163.
29. Wilson, R. W. & Schellman, J. A. (1977) *Biopolymers* 16, 2143–2165.
30. Studier, F. W. (1979) *Virology* 39, 562–574.
31. Rosenberg, A. M. & Studier, F. W. (1969) *Biopolymers* 7, 765–774.
32. Dunn, J. J. & Studier, F. W. (1983) *J. Mol. Biol.* 166, 477–536.
33. Manning, G. S. (1977) *Biophys. Chem.* 7, 95–102.
34. Manning, G. S. (1977) *Biophys. Chem.* 9, 189–192.
35. Manning, G. S. (1978) *Q. Rev. Biophys.* 11, 179–192.
36. Hofrichter, H. J. & Schellman, J. A. (1973) *Jerusalem Symp. Quantum Chem. Biochem.* 5, 787–807.
37. Kemp, J. C. (1969) *J. Opt. Soc. Am.* 59, 95–954.
38. Zimm, B. H. (1956) *J. Chem. Phys.* 24, 269–278.
39. Peterlin, A. & Stuart, J. A. (1939) *Z. Phys.* 112, 1–19.
40. Zimm, B. H. & Crothers, D. M. (1962) *Proc. Natl. Acad. Sci. USA* 48, 905–911.
41. Gill, S. J. & Thompson, D. S. (1967) *Proc. Natl. Acad. Sci. USA* 57, 562–566.
42. Van Holde, K. E. (1971) *Physical Biochemistry*, Hagler, L. & Wold, F., Eds., Prentice-Hall, Englewood Cliffs, N.J.
43. Tschoegl, N. W. (1966) *J. Chem. Phys.* 44, 4615–4617.
44. Kuhn, W. & Grun, F. (1942) *Kolloid Z.* 101, 248–271.
45. Gotlib, Y. & Svetlov, Y. (1966) *Polym. Sci. USSR* 8, 1670–1677.
46. Kramers, H. A. (1946) *J. Chem. Phys.* 14, 415–424.
47. Gotlib, Y. & Svetlov, Y. (1966) *Dokl. Phys. Chem. USSR* 168, 339–342.
48. Shimada, J. & Yamakawa, H. (1976) *Macromolecules* 9, 583–586.
49. Ptitsyn, O. B. & Eizner, Y. E. (1959) *Z. Tekh. Fis.* 39, 1117–1134.
50. Tschoegl, N. W. (1963) *J. Chem. Phys.* 39, 149–153.
51. Tschoegl, N. W. (1964) *J. Chem. Phys.* 40, 473–479.
52. Bloomfield, V. & Zimm, B. H. (1966) *J. Chem. Phys.* 44, 315–343.
53. Harrington, R. E. (1978) *Biopolymers* 17, 919–936.
54. Cairney, K. & Harrington, R. (1982) *Biopolymers* 21, 923–934.
55. Arnott, S. & Hulkins, K. (1972) *Biochem. Biophys. Res. Commun.* 47, 1504–1509.
56. Dougherty, A. M., Causley, G. C. & Johnson, W. C., Jr. (1983) *Proc. Natl. Acad. Sci. USA* 80, 2193–2195.
57. Minyat, E. E., Ivanovn, V. I., Kritzyn, A. M., Minchenkova, L. E. & Schyolkina, A. K. (1978) *J. Mol. Biol.* 128, 397–409.
58. Odijk, T. (1977) *J. Polym. Sci., Polym. Phys. Ed.* 15, 477–483.
59. Odijk, T. & Houwaart, A. C. (1978) *J. Polym. Sci., Polym. Phys. Ed.* 16, 627–639.
60. Odijk, T. (1979) *Biopolymers* 18, 3111–3113.
61. Skolnik, J. & Fixman, M. (1977) *Macromolecules* 10, 944–948.
62. Schurr, J. M. & Allison, S. A. (1981) *Biopolymers* 20, 251–268.
63. Braunlin, W. H., Strick, T. J. & Record, M. T., Jr. (1982) *Biopolymers* 21, 1301–1314.
64. Schellman, J. A. & Parthasarathy, N. (1984) *J. Mol. Biol.* 175, 313–329.
65. Manning, G. S. (1983) *Biopolymers* 22, 689–729.
66. Douthart, R. J. & Bloomfield, V. A. (1968) *Biopolymers* 6, 1297–1309.

Received August 8, 1983
Accepted June 15, 1984