Detection and Characterization of Triple-Helical Pyrimidine-Purine-Pyrimidine Nucleic Acids with Vibrational Circular Dichroism

ARTICLE *in* BIOCHEMISTRY · JULY 1994

Impact Factor: 3.02 · DOI: 10.1021/bi00194a006 · Source: PubMed

3 AUTHORS, INCLUDING:

Petr Pancoska
University of Pittsburgh
101 PUBLICATIONS **3,078** CITATIONS

SEE PROFILE

Timothy A Keiderling
University of Illinois at Chicago
276 PUBLICATIONS **7,305** CITATIONS

SEE PROFILE

Detection and Characterization of Triple-Helical Pyrimidine-Purine-Pyrimidine Nucleic Acids with Vibrational Circular Dichroism

Lijiang Wang, Petr Pancoska,‡ and Timothy A. Keiderling*

Department of Chemistry, University of Illinois at Chicago, Chicago, Illinois 60607

Received December 8, 1993; Revised Manuscript Received May 16, 1994®

ABSTRACT: Vibrational circular dichroism (VCD) spectra were measured in the C=O stretching region for poly(U)*poly(A)·poly(U), poly(dT)*poly(dA)·poly(dT), and poly(U)*poly(dA)·poly(dT). These VCD spectra of the triple-helical structure were dramatically different from those of the corresponding duplexes. The VCD indicates that a very similar base-pair structure is present in these triplexes. The same sign pattern was found for poly(C^+)*poly(I)·poly(C), which implies a generality of structure than can result from the steric constraint of the triple helix conformation. By contrast, the corresponding duplexes are quite different in terms of their VCD. The transitions between triplex, duplex, and single-stranded forms were studied as a function of temperature and interpreted using factor analysis. The relative stabilities of the triplexes lie in the order RNA > DNA > hybrid. Nondegenerate dipole-coupling calculations for a U*A·U oligomer were carried out for the C=O stretching modes to model the spectral changes observed. The experimental absorbance spectra indicate that the bases have nonequivalent H-bonds which can be achieved if a reverse Hoogsteen base-pairing scheme is assumed. The computational VCD results with such a scheme were in better qualitative agreement with experiment than those using the expected Hoogsteen base-pairing scheme.

Triple-helical forms of nucleic acids were discovered more than thirty years ago (Felsenfeld et al., 1957) and have been found to have highly sequence-specific structures. This specificity is exhibited as (a) a pyrimidine-purine-pyrimidine binding motif in which thymine binds to an adenine-thymine base pair (T*A·T)[1] or protonated cytosine binds to a guanine-cytosine base pair (C^+*G·C) (Riley et al., 1966; Thiele and Guschlbauer, 1969) or (b) a purine-purine-pyrimidine motif in which adenine recognizes the adenine-thymine base pair (A*A·T) or guanine recognizes the guanine-cytosine base pair (G*G·C) (Lipsett, 1963; Broitman et al., 1987). Both types of triplexes can occur intermolecularly between polymer–polymer (Guschlbauer, 1976), oligomer–oligomer (Kan et al., 1991; Pilch et al., 1991; Paner et al., 1993), and oligomer–polymer (Strobel et al., 1991). They also can occur intramolecularly within duplex DNA under superhelical stress (Mirkin et al., 1987) or within linear single strands (Haner & Dervan, 1990). The restriction to exclusively pyrimidine or purine strands has been overcome in an intramolecular DNA triplex formed with both PuPuPy and PyPuPy base triplets (Jayasena & Johnston, 1992). Moreover, it was found that a single strand of any sequence can be hybridized to a duplex DNA to form a triplex, if assisted by the RecA protein (Ferrin & Camerini-Otero, 1991). A number of studies on oligonucleotide-directed triple-helix formation have shown that this structure is potentially important in gene mapping (Dervan, 1992) and for chemotherapeutic agents (Riordan & Martin, 1991). Discoveries of repression of transcription *in vitro* by triplex formation (Cooney et al., 1988; Young et al., 1991) and natural triplex binding sites in gene promoters (Durland et al., 1991) indicates that the triplex structure may have importance in biological processes.

The most detailed structural information available for the py-pu-py class of triplexes has been provided by fiber X-ray diffraction data (Arnott et al., 1976; Chandrasekaran & Arnott, 1989). A pyrimidine strand is assigned to lie in the major groove of a Watson–Crick double-stranded helix and form Hoogsteen-type hydrogen bonds with the parallel purine strand. However, in a 31-base intramolecular DNA triplex, most sugar puckers were found to be predominantly C2′-endo (Mayaca et al., 1992) from the analysis of the fine structure of COSY cross peaks. In addition, a different structure involving a reverse Hoogsteen base-pairing scheme was proposed for a pu-pu-py triplex where the third purine-rich strand binds antiparallel to the purine strand of the W–C duplex (Beal & Dervan, 1991). Therefore, the structure of triple-helical nucleic acids and its sequence dependence are far from understood.

Data from several spectroscopic techniques have been used to supplement the low-resolution fiber X-ray data available for triplexes and to aid understanding of this structure in solution or under physiological conditions. Here we will add data from a new technique for triplex characterization. Using a combination of IR and CD, called vibrational circular dichroism (VCD), we demonstrate a new approach to triplex studies and gain possible new insight into its structure.

VCD couples the conformational sensitivity of CD with the spectroscopic advantages of multiple resolved, assignable, and localized transitions inherent to vibrational spectroscopy and has been demonstrated to have significant potential for biopolymer conformational studies (Keiderling, 1993; Keiderling & Pancoska, 1993; Diem, 1991; Freedman & Nafie, 1989). We and others have shown that VCD is indicative of the handedness of both single- and double-stranded polynucleotides using vibrational transitions in both the base deformation and the PO_2^- stretching (1300–1000 cm^{-1}) regions (Annamalai & Keiderling, 1987; Diem, 1991; Wang & Keiderling, 1992, 1993). Theoretical analyses indicate that

* To whom correspondence should be addressed.

‡ Permanent address: Department of Chemical Physics, Faculty of Mathematics and Physics, Charles University, 121 16 Prague 2, Czech Republic.

® Abstract published in *Advance ACS Abstracts,* June 15, 1994.

[1] * denotes Hoogsteen-type bonding between base pairs and · Watson–Crick bonding between base pairs.

0006-2960/94/0433-8428$04.50/0

the VCD in both spectral regions arises primarily from the dipole coupling of the associated oscillators (Gulotta et al., 1989; Wang et al., 1993). In our previous studies of the thermal disproportionation of poly(A)·poly(U) (Yang & Keiderling, 1993), a VCD spectrum was obtained at an intermediate temperature (~52–62 °C) that was distinctively different from those of either single- or double-stranded RNA. The conformer that gave rise to this unique spectrum was then identified by comparative studies to be triplex poly(U)*poly(A)·poly(U). This result drew our attention to the potential of utilizing VCD for identifying triple helices in nucleic acids.

In this paper, we show that VCD (arising primarily from the coupling of transition dipoles) can be employed to detect triple-helical formation as well as to discriminate triplex uniquely from duplex and single strand forms in an unique manner. Three AT(U)-based homopolymer triplexes are studied [poly(U)*poly(A)·poly(U), an RNA; poly(dT)*poly(dA)·poly(dT), a DNA; and poly(U)*poly(dA)·poly(dT), an RNA–DNA hybrid], and the results are compared to the VCD of poly(C^+)*poly(I)·poly(C), a protonated RNA triplex. The conformational transitions between various stranded forms; the relationship between RNA, DNA, and hybrid triplex stabilities; factor analysis of the spectral band shapes to elucidate the polymer conformational state; and dipole-coupling simulation of the experimental IR and VCD spectra are discussed.

MATERIALS AND METHODS

Triplex Preparation. Poly(U)*poly(A)·poly(U) (lot 97F-06661), poly(dA)·poly(dT) (lot 47F03881), poly(dA) (lot 82H6746), poly(dT) (lot 56F6706), poly(U) (lot P-9528), poly(I) (lot P-4154), and poly(C) (lot 63F-4017) were purchased from Sigma and used without further purification. Solution concentrations of single- and double-strand nucleic acids that were used to prepare the triplexes were standardized by measuring their UV absorbance maxima (each occurs near 260 nm). The following molar extinction coefficients were used for the calculations: 8520 for poly(dA), 6000 for poly(dA)·poly(dT), 9350 for poly(U), 6300 for poly(C), and 10 400 for poly(I) (Liquier et al., 1991; Sigler et al., 1962). Triplexes of poly(U)*poly(dA)·poly(dT) and poly(dT)*poly(dA)·poly(dT) were made by combining solutions of duplex poly(dA)·poly(dT) and single-strand poly(dT) or poly(U), respectively, in a 1:1 molar ratio. Solutions of single-strand poly(C) and poly(I) were used in preparing triplex poly(C^+)*poly(I)·poly(C) (C:I = 2:1). All solutions were prepared in D_2O or NaCl/D_2O solution to a final concentration of ~30 mM in phosphate. pH values were measured with a Corning 145 pH meter equipped with a microcombination electrode (Ingold). Under useful VCD concentration conditions, there is about 0.12–0.18 M alkali concentration already present in nonbuffered D_2O solution due to the nucleic acid itself. This is estimated from the difference of polynucleotide concentrations calculated from sample weights and solution volumes and from absorbance measurements. In some cases (as will be noted) more NaCl was added to stabilize the conformation.

Vibrational CD and IR. VCD and IR spectra were measured on the UIC dispersive instrument which has been previously described elsewhere in detail (Keiderling, 1990, 1981). The sample solutions were placed in a cell composed of two CaF_2 windows separated by a 50-μm Teflon spacer. The cell was then fit into a customized variable-temperature cell holder (Yang & Keiderling, 1993; Wang, 1993) whose temperature was referenced to a remote sensor on the cell holder and maintained by a thermostated bath and control unit (Neslab, model RTE-110 and RS-2). The relative temperature is quite stable and resettable with an absolute temperature uncertainty at the sample windows of about 1 °C. To anneal the multistrand forms, samples were preheated to about 85 °C and then slowly cooled down to the desired temperature. The spectra presented in this paper were measured in the base deformation region where C=O stretching modes contribute to the most intense IR transitions for each base except adenine. The C4=C5 mode in A has a significant absorption contributions at about 1625 cm^{-1}.

Dipole Coupling Calculations. Attempts to simulate triplex and duplex VCD spectra in the base deformations were made using the extended coupled oscillator (ECO) model described by Diem and co-workers (Gulotta et al., 1989) which is based on the exciton theory developed earlier by Tinoco (1963) for ECD and later modified by Holzwarth and Chabay (1972) for VCD of dimers. This model has successfully explained the correlation between the sign pattern of VCD and the handedness of some single- and double-stranded nucleic acids (Gullota, 1989; Zhong et al., 1990; Xian et al., 1993; Wang & Keiderling, 1993). To account for the complex triplex VCD spectra, we calculated dipole (D_k) and rotational (R_k) strengths for the kth transition of an N dipole system arising from dipole-coupling interactions between nondegenerate dipoles (Bour & Keiderling, 1993; Wang, 1993). Our formulations are equivalent to those recently reported by Xian et al. (1993), which they termed the nondegenerate ECO or NECO model. These calculations derive from the following relationships:

$$D_k = \sum_{i=1}^{N}\sum_{j=1}^{N} C_{ik}C_{jk}(\vec{\mu}_i \cdot \vec{\mu}_j) \tag{1}$$

$$R_k = -\pi/c \sum_{i=1}^{N}\sum_{j=1}^{N} C_{ik}C_{jk}(\nu_j \vec{T}_j - \nu_i \vec{T}_i)\cdot(\vec{\mu}_i \times \vec{\mu}_j) \tag{2}$$

where c is the speed of light, C_{ik} and C_{jk} are the eigenvector components of the dipole–dipole interaction matrix, $\vec{T}_i$ are the positions of the center of mass of each oscillator expressed as a vector from the coordinate origin, and $\vec{\mu}_i$ are the electric transition dipole moment vectors, whose magnitudes can be estimated to be $D^{1/2}$, where D is the dipole strength of the vibrational transition of the monomer.[2] For these calculations we have assumed that the interaction is dominated by pairwise dipole–dipole coupling between transition dipoles i and j,

$$V_{ij} = \frac{1}{\epsilon_d}\left[\frac{\vec{\mu}_i \cdot \vec{\mu}_j}{|\vec{T}_{ij}|^3} - \frac{3(\vec{\mu}_i \cdot \vec{T}_{ij})(\vec{\mu}_j \cdot \vec{T}_{ij})}{|\vec{T}_{ij}|^5}\right] \tag{3}$$

where ϵ_d is the dielectric constant, ν_i and ν_j are the unperturbed frequencies of the dipoles i and j (in s^{-1}), and $\vec{T}_{ij} = \vec{T}_i - \vec{T}_j$.

The IR absorption spectra of the single-stranded polymers were carefully remeasured here and curve fit to obtain an accurate measure of the band area for determination of the dipole strengths used in our calculations. Concentrations of these single-strand samples were in each case ~0.022 M in phosphate, as determined by UV absorbance. The path length of the spacer used was 0.0049 cm, as calculated from the fringe pattern produced in the transmission of an empty cell. All the coordinates used are from X-ray fiber diffraction data (Chandrasekaran & Arnott, 1989; Arnott et al., 1976).

[2] The D value is obtained by integrating the absorption bands corresponding to the transition of interest in the monomer IR spectrum ($D = (92 \times 10^{-4})\int \epsilon/\nu \, d\nu$ in D^2).

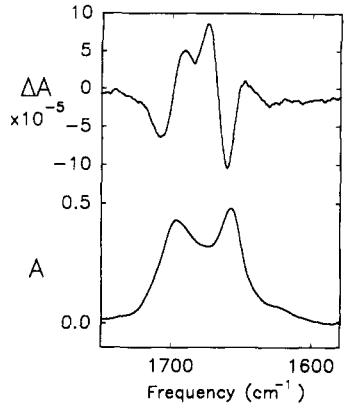

FIGURE 1: VCD (top) and absorbance (bottom) spectra of poly(U)*poly(A)·poly(U) in D_2O for comparison.

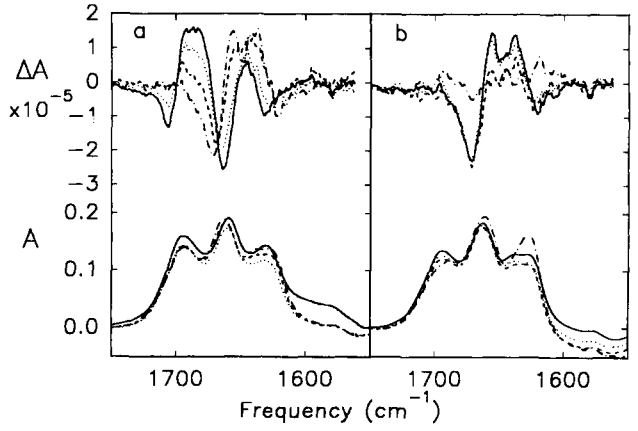

FIGURE 2: VCD and absorbance spectra from the thermal denaturation of poly(dT)*poly(dA)·poly(dT) in D_2O. (a) Triple to double-strand transition: (—) 5 °C; (···) 25 °C; (- - -) 33 °C; (- · - · -) 35 °C. (b) Double to single-strand transition: (—) 40 °C; (· · ·) 60 °C; (- - -) 70 °C; (- · - · -) 72 °C.

RESULTS AND ASSIGNMENTS

In the earlier studies of the thermal disproportionation reaction of poly(A)·poly(U) (U·A) (Yang & Keiderling, 1993), a double- to triple-stranded conformational transition was observed with a T_m at about 53 °C. The triple-helical poly(U)*poly(A)·poly(U) (U*A·U) gave rise to an unusual five-peak (− + − + +) VCD band pattern (Figure 1). In our attempt to carry out a parallel DNA-based experiment, the temperature dependencies of the poly(dA)·poly(dT) (dT·dA) VCD spectra were measured, but no conformational transition to triplex poly(dT)*poly(dA)·poly(dT) (dT*dA·dT) was detected, even if the added NaCl concentration was raised to ~0.30 M, a condition under which U*A·U is well-stabilized (Krakauer & Sturtevant, 1968).

Taking an alternative approach, the VCD of a sample initially prepared as dT*dA·dT was measured. When its temperature was lowered to 5 °C, a totally different spectrum (solid line, Figure 2) from that of duplex dA·dT was observed. If the broad positive VCD band observed at ~1687 cm^{-1} is assumed to encompass two bands, this multiply overlapped spectrum can be viewed as having the same sign pattern (− + + − +) as that of triplex U*A·U (Figure 1), with an additional negative band at ~1632 cm^{-1}. As the temperature was increased to 35 °C, this low-temperature VCD spectrum progressively transformed into a different form (dashed–dotted spectra in Figure 2a). The intermediate band shape is dominated by a positive couplet correlated to the central absorption feature and an overlapped positive VCD to lower frequency and is virtually identical to the VCD band shape

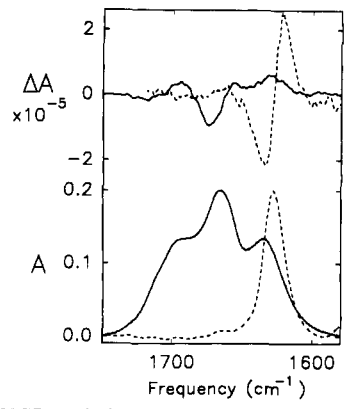

FIGURE 3: VCD and absorption spectra of poly(dT) (—) and poly(dA) (- - -) in D_2O at 25 °C for comparison to Figure 2.

of duplex dA·dT (Wang, 1993; Zhong et al., 1990). The temperature-dependent spectra in Figure 2a have an isobestic point in absorption that coincides with the isodichroic point in VCD at 1668 cm^{-1}, both observations being consistent with the sample having undergone a two-state transition. The intermediate duplex spectrum then transformed into a very weak VCD by 72 °C (dashed–dotted line, Figure 2b). It is worth noting that the absorption spectral change has a much smaller variation than does the VCD spectrum for both of these conformational transitions (Figure 2a and b).

In order to facilitate assignment of these VCD spectra for multistranded forms, it is necessary to establish the characteristics of the single-stranded spectra. In Figure 3, the VCD and absorption spectra are compared for poly(dT) (solid) and poly(dA) (dashed) at 25 °C under the same conditions as for the dT*dA·dT experiment. Spectra for both single-stranded DNAs were also measured at 5 and 72 °C and gave similar band shapes to those shown in Figure 2 but lost roughly half their intensity as temperature was increased. The poly(dT) VCD spectrum is quite weak (note increase in absorbance as compared to Figure 2) but resembles the duplex spectrum in that a negative VCD feature occurs to the high-frequency-energy side of the central maximum-intensity absorption feature at 1665 cm^{-1}. Weaker positive features occur to either side of this band. By comparison, the poly(dA) VCD is much simpler. A fairly symmetric positive couplet is centered over the single strong absorption band at 1627 cm^{-1}. We can eliminate the possibility of our having detected just a mixed state in the low-temperature dT*dA·dT VCD by trying to computationally synthesize the multistrand spectra from reasonable combinations of the dA·dT duplex and dA and dT single-strand VCD spectra. The low temperature VCD observed (Figure 2a) can be solely assigned to triple-helical dT*dA·dT on the basis of this test and its similarity to the spectra of its RNA analog, U*A·U.

The same sort of experiment was carried out for a 1:1 stoichiometric mixture of poly(dA)·poly(dT) with poly(U), but under low salt conditions, very little triplex was formed. When 0.1 M NaCl was added, a spectrum almost identical to that of dT*dA·dT was obtained at 5 °C (solid line, Figure 4). The VCD spectrum of duplex dA·dT was fully recovered at 40 °C (dashed line, Figure 4), a slightly higher temperature than found for dT*dA·dT. The observed spectral change with increasing temperature is in reasonable agreement with the triplex to duplex transition data in Figure 2a. The parallels in band shape and behavior confirm that a triple-helix form of poly(U)*poly(dA)·poly(dT) (U*dA·dT) is stable at low temperature with added salt and has a very similar conformation to that of dT*dA·dT. In both of these triplexes, as was

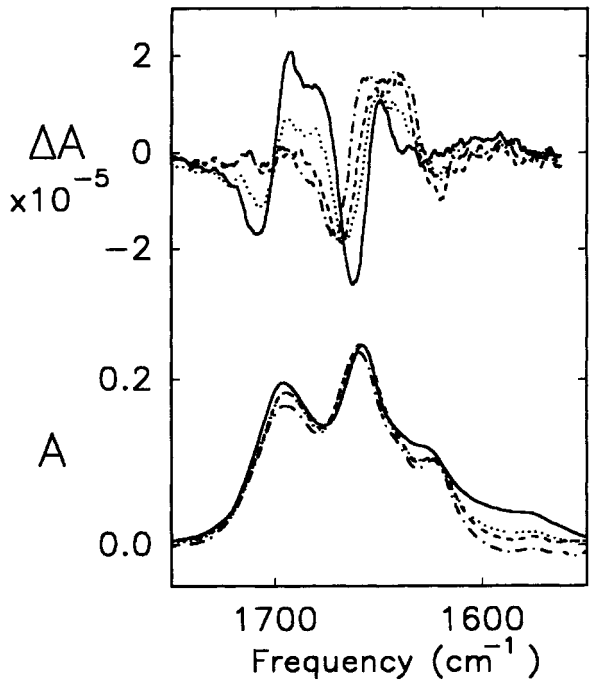

FIGURE 4: VCD and absorbance spectra for the triple- to double-strand transition of poly(U)*poly(dA)·poly(dT) occurring upon thermal denaturation in 0.1 M NaCl/D_2O: (—) 5 °C; (···) 25 °C; (- - -) 30 °C; (-·-·-) 40 °C.

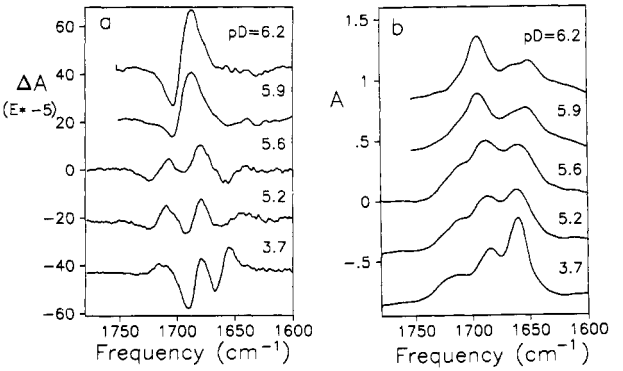

FIGURE 5: VCD (a) and absorption (b) spectra of poly(C)*poly(I)·poly(C) in D_2O at pD 6.2, 5.9, 5.6, 5.2, and 3.7.

seen for the U*A·U triplex, there is a significant spectral difference from the duplex. A high-frequency (~1710 cm^{-1}) positive VCD couplet is seen for each triplex while little if any VCD occurs for these higher energy transitions in the duplex. The IR changes are quite minor by comparison to the extensive overall band-shape changes seen for the triplex–duplex transition in all three of these systems.

To compare with AT(U)-based triplexes, VCD and IR absorption spectra of a poly(I)·poly(C) sample prepared at I:C = 1:2 (CIC) were studied at various pD values. The results are given in FIgure 5a and b for VCD and absorption, respectively. The VCD seen at pD 6.2 for the CIC system (Figure 5a) has the same band shape as A-form poly(I) (Annamalai & Keiderling, 1987), being dominated by a positive couplet centered over the most intense absorbance at ~1697 cm^{-1}. When the pD was lowered to 5.6, this simple couplet VCD collapsed into a weaker and more structured VCD spectrum with a (− + + − +) sign pattern. This pD 5.6 transformed five-feature VCD pattern is attributable to the triplex C$^+$*I·C conformation.[3] The high-frequency absorbance band at about 1712 cm^{-1} indicates protonation of cytosine at N3. Further decrease of pD led to the disappearance of the

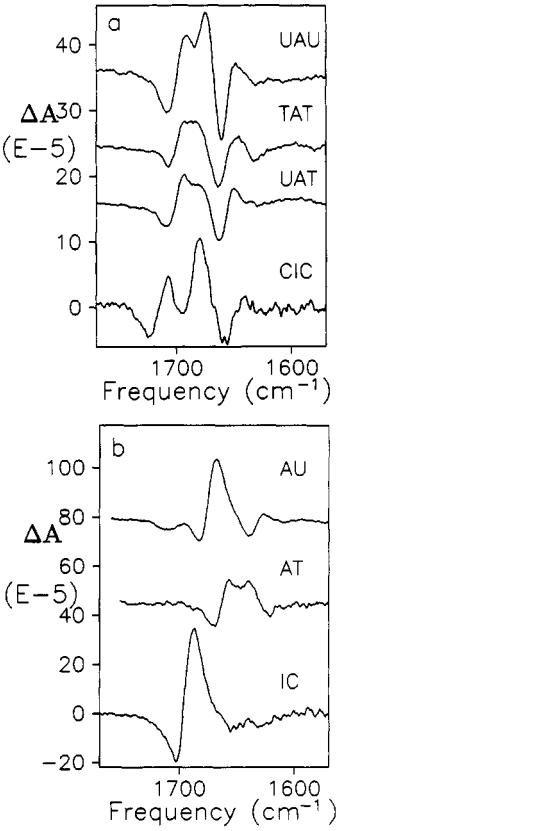

FIGURE 6: Overlapped VCD spectra of (a) the triplex nucleic acids (top to bottom) U*A·U (in D_2O, 25 °C), dT*dA·dT (in D_2O, 5 °C), U*dT·dA (in 0.1 M NaCl/D_2O, 5 °C), and C$^+$*I·C (in D_2O at pD 5.6, 25 °C) and (b) the duplex nucleic acids A·U (in 0.1 M NaCl/D_2O, 25 °C), dA·dT (in D_2O, 25 °C), and I·C (in D_2O at pD 5.6, 25 °C).

negative 1720-cm^{-1} VCD band and a sharpening and increase in intensity of the bands below 1712 cm^{-1}.

DISCUSSION AND ANALYSIS

VCD Indicates Common Structure for Triple Helix. When viewed together, our data indicate that VCD provides a sensitive means of detecting triplex formation. In Figure 6a, the VCD spectra of the four homopolymer triplexes we studied are stacked over each other for the sake of comparison. From high to low frequency, the VCD spectra in the C=O stretching base deformation region of U*A·U, dT*dA·dT, and U*dA·dT have the same (− + + − +) band sign pattern and are very similarly shaped. Due to its quite different molecular structure, the triplex formed between I and C has different absorption frequencies than found in the AT(U) system. This naturally leads to a difference in the observed VCD band shape which is largely determined by spectral overlap of multiple polymer transitions of varying sign and magnitude; however, the characteristic five-peak (− + + − +) pattern is still present in the C$^+$*I·C VCD spectrum.

On the other hand, the duplexes from which these triplexes are derived, poly(A)·poly(U), poly(dA)·poly(dT), and poly(I)·poly(C), have very different VCD patterns for the C=O stretching modes as compared in Figure 6b. These differences in the duplex VCD spectra, as contrasted with the similarities of the corresponding triplex VCD, indicate that the triplexes must share a very similar base-pair structure and stacking pattern.

[3] ECD of this sample is consistent with the C$^+$*I·C ECD reported by Thiele and Guschlbauer (1969).

The close comparison of the VCD spectra for different sequences eliminates alternate explanations for the observed VCD patterns. For example, poly(U) is known to adopt a double-strand conformation at 5 °C due to aggregation (Saenger, 1984). This double-stranded structure cannot be the source of the U*dA·dT triplex six-peak band pattern because, first, the characteristic high-energy band is not present in low-temperature poly(U), second, U*dA·dT and dT*dA·dT have very similar spectra to U*A·U, and, third, the U*A·U triplex spectrum is stable at high temperatures; therefore, its common elements with the other AT(U) VCD spectra cannot result from duplex U, which is only stable as a low-temperature species.

Most single- and double-stranded nucleic acids studied so far yielded a characteristic dominant positive VCD couplet over the most intense IR absorption band for a right-handed helical structure and a negative one for a left-handed structure (Wang & Keiderling, 1993). By contrast, the more complex five-peak (− + + − +) pattern found in the triplex spectra reported here is not only characteristic of the triple helix but also uniquely discriminates the triplex VCD spectra from all the duplex and single-strand nucleic acid VCD spectra we have studied.

The difference between triplex and duplex IR absorption spectra is much less useful in distinguishing these forms than is the variance found in their VCD spectra, at least for the AT(U) systems. Claims that one can use FTIR absorbance spectra to discriminate triplexes from duplexes (Liquier et al., 1991) on the basis of there only being two major C=O stretching features for AT(U) triplexes but more for duplexes clearly fall short when a DNA such as dT*dA·dT is considered (Figure 2a). Yet the VCD patterns for all the AT(U) triplexes we have studied (RNA, DNA, hybrid) are exceptionally consistent. That they also agree in form with the VCD spectra of the C^+*I·C triplex is particularly satisfying. This again demonstrates that VCD is more sensitive to the three-dimensional conformation than is absorbance. That this should be expected follows from the chiral sensitivity of VCD (Keiderling & Pancoska, 1993).

Many triple helices have been formed between double-stranded polynucleotides and oligonucleotides. Since VCD is primarily dependent on short-range interactions (Dukor & Keiderling, 1991), these patterns are expected to hold for oligomer triplexes also. In addition, oligomer–polymer triplex formation could be detected if the triplex component were a significant fraction of the total nucleic acid concentration.

Factor Analysis Confirms Biphasic Transitions. Factor analysis (Pancoska et al., 1979; Malinowski & Howery, 1980) is a method of decomposing a set of spectra into linear combinations of subspectra. The coefficients of each subspectrum represent its contribution to the experimental spectrum and, by inference, the contribution of the structural components represented by that subspectrum to the total structure. Ideally, for a study of conformational variation with environmental change such as temperature or pH, a plot of coefficient values versus the variable that induces the structural change should reflect the change of the structural component correlated to the observed spectral variation.

Factor analysis was performed on a set of nine spectra for the VCD of the dT*dA·dT samples. For these data three important subspectra were identified which are significantly above the noise level. Coefficients of the 1st and 2nd subspectra vary with temperature (Figure 7a) in such a way that a biphasic transition can be easily identified with two T_m, one at ~30 °C and another at ~71 °C.

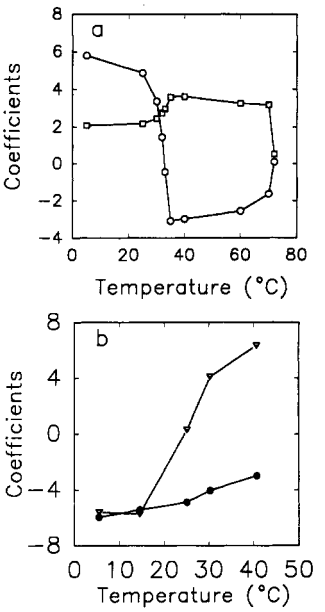

FIGURE 7: Variation versus temperature of VCD subspectrum coefficients as derived from factor analysis. (a) dT*dA·dT: (□) subspectrum 1; (○) subspectrum 2. (b) U*dA·dT: (●) subspectrum 1; (▽) subspectrum 2.

For the U*dA·dT set of data, factor analysis was carried out on the spectra representing only the triplex to duplex transition. The duplex to single-strand transition is the same as that seen in the dT*dA·dT data set (above), since it is also involves the transition from duplex dA·dT to single-strand poly(dA) and poly(dT).[4] The coefficient versus temperature curves for U*dA·dT are given for the first and second subspectra in Figure 7b. The change with temperature is more dramatic for the coefficient of the second subspectrum than for that for the first in this case. Both are consistent with a cooperative two-state transition from the triplex to duplex forms, but different profiles and transition temperatures result for each, as is discussed more below. A result consistent with this was also obtained for the U*A·U data measured earlier (Yang & Keiderling, 1993), which was reported in terms of ΔA variation with temperature. In our view, the factor analysis decomposition of the spectra was very useful in determining that the transition from triplex to duplex was actually biphasic and in locating the transition temperatures for the U*dA·dT and dT*dA·dT systems, which were not easily seen using just ΔA variation with temperature.

Comparison of Relative Stabilities. When measured under the same experimental conditions (appropriate for VCD study), our results show that the all-RNA triplex, U*A·U, is stable to ~57 °C (Yang & Keiderling, 1993), the all-DNA triplex, dT*dA·dT, is disproportionated at a much lower temperature (T_m ~ 30 °C), and the hybrid, U*dA·dT, can only be stabilized at higher total salt concentration (>0.2 M) and low temperature (<15 °C). This implies that the relative stability of these triplexes lies in the order RNA > DNA > hybrid. This is in qualitative agreement with the stability ordering of the same polynucleotide triplexes based on their melting temperatures observed by Riley et al. (1966) but not with that of the GC-dominant 16-mer triplexes based on temperature-dependent CD determinations of free energies (Roberts and Crothers, 1992). By contrast, the duplexes dA·dT and A·U

[4] That U·dA is not formed is evident from the 40 °C band shape (Figure 4) and is consistent with the relative stabilities of the DNA and RNA duplexes (Guschlbauer, 1976).

from which these triplexes derive have the opposite relative stability ordering (Guschlbauer, 1976). It has been suggested that the extra hydroxyl at C2′ in the RNA interacts with solvent so that a stronger solvent–nucleotide interaction is possible for RNA versus DNA triplexes (Riley et al., 1966). The fact that increasing ionic strength stabilizes the triplex form ($\Delta T/\Delta\log[Na^+] = 30$ °C) at a faster rate of change than the duplex ($\Delta T/\Delta\log[Na^+] = 20$ °C) (Guschlbauer, 1976) may be a consequence of there being three short phosphate–phosphate interactions in the triplex and only one in the duplex. These previous observations of triplex versus duplex stabilities for model DNAs and RNAs are also consistent with the fact that we could not obtain dT*dA·dT from disproportionation of poly(dA)·poly(dT) as we did in the case of U*A·U. Therefore, our data offer further confirmation of previous observations that composition is an important parameter in determining the stability of triplex formation.

Implying a strong structural similarity, these AT(U) triplexes (Figure 6a) have more similar VCD spectra than do the duplexes dA·dT and A·U (Figure 6b). The RNA and DNA duplexes do have distinct conformations, but in most DNA systems, the change from B– to A-form has only a small effect on the VCD band shape (Wang & Keiderling, 1992). Even the VCD spectrum of triplex C⁺*I·C has the same band-shape pattern, accounting for frequency shifts, as that of the AT(U) triplexes while duplex I·C has a very different VCD band shape. Recent VCD measurements in our laboratory (Huang & Keiderling, unpublished results) showed a consistent five-peak (– + + – +) VCD pattern for another triplex, C⁺*G·C, which adds proof that this VCD pattern is indeed characteristic of triplexes. The stability and other properties of the protonated G(I)C-based triplexes will be discussed separately.

Identical base pairing for AT(U) triplexes was proposed by Liquier et al. (1991) on the basis of their having almost identical IR absorbance spectra. The same claim is made for C⁺*I·C and dC⁺*dG·dC by Akhebat et al. (1992). The IR spectra of the two (AT/GC) py-pu-py triplex families are quite different, but their VCD spectra still have strong similarities. Our consistent VCD data thus suggest that both types of triplexes have identical base stacking patterns. The electrophoretic mobility of a series of mixed-sequence 16-mer triplexes (Roberts & Crothers, 1992) has been measured, and the dispersion of these values as compared to their duplex analogs has been interpreted to indicate that the triplexes have a more uniform structure than do the duplexes. All these phenomena support the suggestion that there is great similarity between triplexes. The ultimate cause of this consistency may be the tightly packed structure in the triplex form, which may favor a rigid, less variable polymeric conformation, especially in terms of the base stacking.

Our data do not eliminate the possibility of small variations in the backbone structure, as has been proposed on the basis of NMR (Macaya et al., 1992), FTIR (Akhebat et al., 1992), and electrophoretic mobility experiments (Roberts & Crothers, 1992) with hybrid triplexes. Because the vibrational modes of the ribose phosphate groups appear significantly below 1500 cm^{-1} and do not significantly couple with the base deformation modes, the VCD data shown here, that arise from the base vibrations, will only reflect variation in the backbone stereochemistry in so far as it affects the base stacking pattern.

Analysis of NECO Calculations. The coupled oscillator model can qualitatively predict the sign pattern of the dominant VCD spectral features for several different duplex forms of

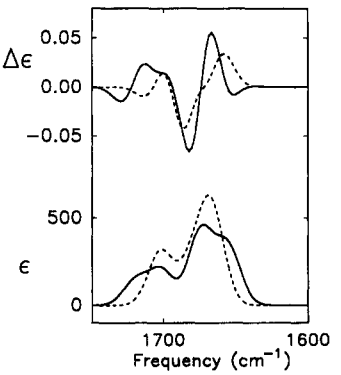

FIGURE 8: Comparison of NECO calculated spectra of $(U)_{12}$*$(A)_{12}$·$(U)_{12}$ with W–C + Hoogsteen base pairs (- - -) to those of $(U)_{12}$*$(A)_{12}$·$(U)_{12}$ with W–C + reverse Hoogsteen base pairs (—). Dipole/dipole strength (D^2)/frequency (cm^{-1}): C2=O(W–C and Hoogsteen)/0.116/1692; C4=O(W–C and Hoogsteen)/0.1444/1669; C2=O(reverse Hoogsteen)/0.116/1707; C4=O(reverse Hoogsteen)/0.144/1656. $\epsilon_d = 2$. Spectra are normalized to per dipole and simulated with a Gaussian band shape of half width 12 cm^{-1}.

DNA and RNA. But for the triplex, it should be clear that the multiband VCD spectrum shown here (Figure 6a) prohibits a simple interpretation based on the assumptions of the dipole-coupling mechanism. To bear this out, it is not surprising that our efforts to simulate the complex triplex VCD spectrum failed with a calculation based on the simple ECO model which considered only coupling of degenerate oscillators. To account for the complexity, we next undertook a series of NECO model calculations to determine if the VCD band shape was at all consistent with this level of semiclassical theoretical analysis. For simplicity, all comparisons will be made to the U*A·U VCD results. The C2=O (U), C4=O (U), and C4=C5 (A) stretching modes are considered to dominate the IR spectra in the base deformation region of the AU-containing species. For simplicity, the transition dipole moments are taken to lie along those double bonds. Our initial calculations considered the transition dipoles in the monomers U and A to be unperturbed and thus used dipole strengths and frequencies derived from their IR spectra for the calculations. The results were not a significant improvement over the ECO simulation noted above. We then used dipole strengths taken from poly(A)·poly(U) for C2=O and C4=O of U and ignored the C4=C5 from poly(A) in the following calculations. The effect of the A mode is expected to be minimal on the basis of our preliminary computational tests and its very small dipolar strength in the U*A·U triplex (Liquier, et al., 1991; Yang & Keiderling, 1993).

For a $(U)_{12}$ third strand that was Hoogsteen base-paired to an $(A)_{12}$·$(U)_{12}$ Watson–Crick (W–C) duplex, the calculated (U)*(A)·(U) spectra using the NECO model are shown in Figure 8 (dashed line). While the simulated absorption spectrum is reasonably consistent with the experimental band shape, the computed VCD pattern fails to match the five-peak (– + + – +) pattern observed experimentally for (U)*(A)·(U) (Figure 1). The fairly regular VCD pattern obtained results from assigning a uniform frequency and dipole strength to each of the C=O groups in both U bases. Such near degeneracy is expected, since the C4=O groups are H-bonded and the C2=O are not for each poly(U) strand whether under Watson–Crick or Hoogsteen base pairing.

However, the experimental IR absorption spectrum is not compatible with equivalent C=O groups in the two U-bases. The FTIR spectrum of (U)*(A)·(U) at higher resolution (2 cm^{-1}, Figure 9, thick line, bottom) shows three absorption bands. The band at 1660 cm^{-1} has the same frequency as

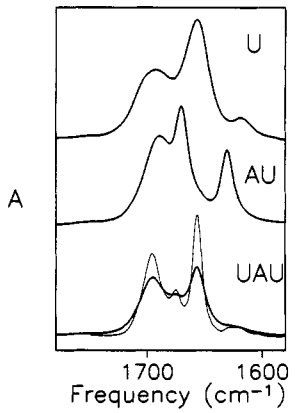

FIGURE 9: Comparison of IR spectra of poly(U), poly(A)·poly(U), and poly(U)*poly(A)·poly(U): top, poly(U) at resolution = 4 cm^{-1} in D_2O; middle, poly(A)·poly(U) at resolution = 4 cm^{-1} in D_2O; bottom, original (thick line) and Fourier self-deconvolved FTIR spectrum (thin line) of poly(U)*poly(A)·poly(U) at resolution = 2 cm^{-1} in D_2O. Deconvolution was performed on an FTS 40 spectrometer with a half width of 14 cm^{-1}, *k* factor of 2.2, and Bessel apodization function.

seen for the free C4=O which is not H-bonded in poly(U) (Figure 9, top), but the structural model using a Hoogsteen base-paired third strand does not allow a free C4=O group. A new band at about 1675 cm^{-1} is easily identifiable in the deconvolved FTIR spectrum of (U)*(A)·(U) (Figure 9, thin line, bottom). This is very close in frequency to the H-bonded C4=O stretching band seen in (A)·(U) (Figure 9, middle) and would correspond to that expected for the Watson–Crick H-bonded U-strand in the triplex. Finally, the broad band at ~1695 cm^{-1} in (U)*(A)·(U) is higher in frequency than the C2=O stretch in (A)·(U), which would be consistent with their being both free and H-bonded C2=O groups in the triplex structure. Such observations were, in fact, made in the early triplex literature (Miles, 1964) but have apparently been ignored in the interim period.

Such an inequivalent arrangement of C=O groups can be achieved by using reverse Hoogsteen pairing to model the triplex structure. This would result in the C4=O of the U-strand in the Watson–Crick duplex and the C2=O in the third U-strand being H-bonded while the other two carbonyls would be free. The four C=O groups in such a structure have different environments and different frequencies and become nonequivalent, which leads to their reflecting the observed properties of the experimental IR absorption spectrum. The result of an NECO calcultion based on the geometrical parameters derived from a U*A·U triplex configured with reverse Hoogsteen base pairs (solid line) is compared in Figure 8 with the one for Hoogsteen base pairs (dashed line) discussed previously.

As can be seen in Figure 8, when the third strand is switched from a Hoogsteen to a reverse Hoogsteen base-pairing structure, the characteristic triplex VCD sign pattern (− + + − +) emerges. Of particular note is the characteristic triplex high-energy positive couplet followed by an overlapped positive band that is predicted with this reverse Hoogsteen model. However, the experimental (U)*(A)·(U) absorption spectrum was better predicted for the Hoogsteen base-pair structure, due to the spread in frequencies being overestimated in the reverse Hoogsteen case. This latter situation can be remedied by imposing an assumed dielectric constant on the interaction computation (eq 3).

The reverse Hoogsteen base-pairing structure coupled with the dipole-coupling mechanism gives a physically based accounting for the VCD pattern for the U*A·U triplex. It supports the concept that all the C=O dipoles should be nonequivalent in order to explain the experimentally observed IR absorption spectrum and to produce the complex VCD pattern seen for the triplex. This nonequivalence, in turn, provides a means for the dipole-coupling model to act as a major source of the observed triplex nucleic acid VCD just as it apparently does for duplexes. Due to lack of other experimental evidence for such a base-pairing structure, we cannot propose that this structure is general for py-pu-py triplexes, but our calculation certainly raises the possibility that the reverse Hoogsteen structure should be considered. Whatever the structure is, our data do indicate that it should be quite uniform for these triplexes. More work needs to be done in refining the geometry, dipole frequencies and dipole strengths, dielectric constant, and other parameters used in these calculations to confirm the structural implications of this work.

Since there is evidence of the existence of reverse Hoogsteen base-pair structure in the pu-pu-py type of triplex and in triplexes involving α-anomeric ribose units (Taillandier, 1993), comparison between the VCD of py-pu-py and pu-pu-py triplexes could certainly shed light on the viability of the reverse Hoogsteen model employed here. Work is underway in our laboratory to address this problem. Similarly, the parallel of the AT(U) and G(I)C VCD for triplex forms suggests that NECO calculations on both Hoogsteen and reverse Hoogsteen base-pair structures for the G(I)C triplexes which have more isolated transition dipoles would be useful for comparison. Such studies are also planned.

ACKNOWLEDGMENT

This work was supported by a grant from the National Institutes of Health (GM 30147). We are grateful to Prof. A. S. Benight for discussion of preliminary results and Dr. V. Baumruk for his help with preparing factor analysis results and figures.

REFERENCES

Akhebat, A., Dagneaux, C., Liquier, J., & Taillandier, E. (1992) *J. Biomol. Struct. Dyn.* 10, 577–588.

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

Arnott, S., Bond, P. J., Selsing, E., & Smith, P. J. C. (1976) *Nucleic Acids Res.* 3, 24559–24570.

Beal, P. A., & Dervan, P. B. (1991) *Science* 251, 1360–1363.

Bour, P., & Keiderling, T. A. (1993) *J. Am. Chem. Soc. 115*, 9602–9607.

Broitman, S. L., Im, D. D., & Fresco, J. R. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5120–5124.

Chandrasekaran, R., & Arnott, S. (1989) in *Landolt–Bornstein New Series* (Saenger, W., Ed.) Vol. VII (1b), p 104. Springer-Verlag, Berlin.

Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J., & Hogan, M. E. (1988) *Science 241*, 456–459.

Dervan, P. B. (1992) *Nature 359*, 87–88.

Diem, M. (1991) *Proc. SPIE—Int. Soc. Opt. Eng. 1432*, 28–36.

Dukor, R. K., & Keiderling, T. A. (1991) *Biopolymers 31*, 1747–1761.

Durland, R. H., Kessler, D. J., Gunnell, S., Duvic, M., Pettitt, B. M., & Hogan, M. E. (1991) *Biochemistry 30*, 9246–9255.

Felsenfeld, G., Davies, D. R., & Rich, A. (1957) *J. Am. Chem. Soc. 79*, 2023–2024.

Ferrin, J., & Camerini-Otero, D. R. (1991) *Science 254*, 1494–1497.

Freedman, T. B., & Nafie, L. A. (1989) *Proc. SPIE—Int. Soc. Opt. Eng. 1057*, 15–25.

Gulotta, M., Goss, D. J., & Diem, M. (1989) *Biopolymers 28*, 2047–2058.
Guschlbauer, W. (1976) *Nucleic Acid Structure*, Springer-Verlag, New York.
Haner, R., & Dervan, P. B. (1990) *Biochemistry 29*, 9761–9765.
Holzwarth, G., & Chabay, I. (1972) *J. Chem. Phys. 57*, 1632–1635.
Jayasena, S. D., & Johnson, B. H. (1992) *Biochemistry 31*, 320–327.
Kan, L. S., Callahai, D. E., Trapane, T. L., Miller, P. S., Ts'o, P. O. P. & Huang, D. H. (1991) *J. Biomol. Struct. Dyn. 5*, 911–933.
Keiderling, T. A. (1981) *Appl. Spectrosc. Rev. 17*, 189–226.
Keiderling, T. A. (1990) in *Practical Fourier Transform Infrared Spectroscopy* (Krishnan, K., and Ferraro, J. R., Eds.) pp 203–284, Academic Press, San Diego.
Keiderling, T. A. (1993) in *Physical Chemistry of Food Processes* (Baianu, I. C., Pessen, H., & Kumosinski, T. F., Eds.) Vol. 2, pp 307–336, Van Nostrand Reinhold, New York.
Keiderling, T. A., & Pancoska, P. (1993) in *Biomolecular Spectroscopy* (Clark, R. J. H., & Hester, R. E., Eds.) pp 267–315. John Wiley, Chichester, U.K.
Krakauer, H., & Sturtevant, J. (1968) *Biopolymers 6*, 491.
Lipsett, M. N. (1963) *Biochem. Biophys. Res. Commun. 11*, 224–228.
Liquier, J., Coffinier, P., Firon, M., & Taillandier, E. (1991) *J. Biomol. Struct. Dyn. 3*, 437–445.
Malinowski, E. R., & Howery, D. G. (1980) *Factor Analysis in Chemistry*, Wiley, New York.
Mayaca, R. F., Schultze, P., & Feigon, J. (1992) *J. Am. Chem. Soc. 114*, 781–783.
Miles, H. T. (1964) *Proc. Natl. Acad. Sci. U.S.A. 51*, 1104.
Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. M., Filippov, S. A., & Frank-Kamenetskii, M. (1987) *Nature 330*, 495–497.
Pancoska, P., Fric, I. & Blaha, K. (1979) *Collect. Czech. Chem. Commun. 44*, 1296–1312.
Paner, T. M., Gallo, F. J., Doktycz, M. J., & Benight, A. S. (1993) *Biopolymers* (in press).
Pilch, D. S., Levenson, C., & Shafer, R. H. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 1924–1946.
Pilch, D. S., Levenson, C., & Shafer, R. H. (1991) *Biochemistry 30*, 6081–6087.
Riley, M., Maling, B., & Chamberlin, M. J. (1966) *J. Mol. Biol. 20*, 359–389.
Riordan, M. L., & Martin, J. C. (1991) *Nature 350*, 442.
Roberts, R. W., & Crothers, D. M. (1992) *Science 258*, 1463–1466.
Saenger, W. (1984) in *Principles of Nucleic Acid Structure*, pp 311–313, Springer-Verlag, New York.
Sigler, P. B., Davies, D. R., & Miles, H. T. (1962) *J. Mol. Biol. 5*, 709.
Strobel, S. A., Doucette-Stamm, L. A., Riba, L., Housman, D. E., & Dervan, P. B. (1991) *Science 254*, 1639–1642.
Taillandier, E. (1993) in *Proceedings of FEBS Advanced Course on FT-IR of Biomolecules*, Robert Koch Institut des Bundesgesundheitsamts, Berlin.
Thomas, G. A., & Peticolas, W. L. (1983) *J. Am. Chem. Soc. 105*, 993–996.
Tinoco, I. (1963) *Radiat. Res. 20*, 133–139.
Wang, L. (1993) Ph.D. Thesis, University of Illinois at Chicago.
Wang, L., & Keiderling, T. A. (1992) *Biochemistry 31*, 10265–10271.
Wang, L., & Keiderling, T. A. (1993) *Nucleic Acids Res. 21*, 4127–4232.
Wang, L., Yang, L., & Keiderling, T. A. (1994) (submitted).
Xian, T., Goss, D. J., & Diem, M. (1993) *Biophys. J. 65*, 1255.
Yang, L., & Keiderling, T. A. (1993) *Biopolymers 33*, 315–327.
Young, S. L., Krawaczyk, S. H., Matteucci, M. D., & Toole, J. J. (1991) *Proc. Natl. Acad. Sci. 88*, 10023–10026.
Zhong, W., Gulotta, M., Goss, D. J., & Diem, M. (1990) *Biopolymers 29*, 7485–7491.