Thermodynamic, Spectroscopic, and Equilibrium Binding Studies of DNA Sequence Context Effects in Four 40 Base Pair Deoxyoligonucleotides

Peter M. Vallone‡ and Albert S. Benight*

Department of Chemistry, 845 West Taylor Street, Room 4500, University of Illinois, Chicago, Illinois 60607

Received February 10, 2000; Revised Manuscript Received April 17, 2000

ABSTRACT: Effects of different end sequences on melting, circular dichroism spectra (CD), and enzyme binding properties were investigated for four 40 base pair, non-self-complementary duplex DNA oligomers. The center sequences of these oligoduplexes have either of two 22 base pair modules flanked on both sides by sequences differing in AT content. Temperature-induced melting transitions monitored by differential scanning calorimetry (DSC) and ultraviolet absorbance were measured for the six duplexes in buffered 115 mM Na^+ solutions. Values of the melting transition enthalpy, ΔH_{cal}, and entropy, ΔS_{cal}, were obtained directly from DSC experiments. Melting transition parameters, ΔH_{vH} and ΔS_{vH}, were also estimated from a van't Hoff analysis of optical melting curves collected as a function of DNA concentration, assuming that the melting transition is two-state. Melting free energies (20 °C) evaluated from DSC melting experiments on the four duplex DNAs ranged from −52.2 to −77.5 kcal/mol. Free energies based on the van't Hoff analysis were −37.9 to −58.8 kcal/mol. Although the values are different, trends in the melting free energies of the four duplex DNAs as a function of sequence were identical in both DSC and optical analyses. Subject to several assumptions, values for the initiation free energy were estimated for each duplex, defined as $\Delta G_{int} = \Delta G_{cal} - \Delta G_{pred}$, where ΔG_{cal} is the experimental free energy at 20 °C determined from the experimentially measured values of the transition enthalpy, ΔH_{cal}, and entropy, ΔS_{cal}. The predicted free energy of the sequence, ΔG_{pred}(20 °C), is obtained using published nearest-neighbor sequence stability values. For three of the four duplexes, values of ΔG_{int} are essentially nil. In contrast, the duplex with 81.8% GC has a considerably higher estimate of $\Delta G_{int} = 7.1$ kcal/mol. The CD spectra for the six duplexes collected over the wavelength range from 200 to 320 nm are also sequence-dependent. Factor analysis of the CD spectra by singular value decomposition revealed that the experimental CD spectra could be reconstructed from linear combinations of two minor and one major subspectra. Changes in the coefficients of the major subspectrum for different sequences reflect incremental sequence-dependent variations of the CD spectra. Equilibrium binding by *Bam*HI restriction endonuclease to the 40 base pair DNAs whose central eight base pairs contain the recognition sequence for *Bam*HI restriction enzyme bounded by A·T base pairs, 5′-A-*GGATCC*-A-3′ was investigated. Binding assays were performed by titering *Bam*HI against a constant concentration of each of the duplex DNA substrates, in the absence of Mg^{2+}, followed by analysis by gel retardation. Under the conditions employed, the enzyme binds but does not cleave the DNAs. Results of the assays revealed two binding modes with retarded gel mobilities. Binding isotherms for the fraction of bound DNA species versus enzyme concentration for each binding mode were constructed and analyzed with a simple two-step equilibrium binding model. This analysis provided semiquantitative estimates on the equilibrium binding constants for *Bam*HI to the four DNAs. Values obtained for the binding constants varied only 7-fold and ranged from 6×10^{-8} to 42×10^{-8} M, with binding free energies from −8.6 to −9.7 (± 0.2) kcal/mol depending on the sequence that flanks the enzyme binding site. Unlike what was found earlier in binding studies of the 22 base pair duplexes that constitute the core modules of the present 40-mers [Riccelli, P. V., Vallone, P. M., Kashin, I. , Faldasz, B. D., Lane, M. J., and Benight, A. S. (1999) *Biochemistry 38*, 11197−11208], no obvious relationship between binding and stability was found for these longer DNAs. Apparently, effects of flanking sequence stability on restriction enzyme binding may only be measurable in very short duplex deoxyoligonucleoties.

A number of reports have found that cleavage by several restriction enzymes at specific sequence "cognate" sites on duplex DNA can vary significantly depending on the context of sequences flanking the cognate sites (*1*−*5*). Although such context effects appear to be rather generic, the molecular and thermodynamic basis of their origins remains undefined. Binding studies of DNA oligomers having a single preferable cognate binding site provide a direct means with which to dissect and evaluate subtle effects of different flanking sequence environments on restriction enzyme and other ligand binding (*1*, *2*, *6*, *7*). Our studies of very short duplex DNA oligomer systems (12−22 base pairs) have suggested that the thermodynamic stability of flanking sequences might provide at least a partial explanation of such context effects.

Several years ago we reported results of measurements of rates of first strand cleavage by *Alu*I restriction enzyme at its cognate sequence 5′-A-G−C-T-3′, imbedded in short duplex DNA oligomers. In these DNAs, 12−20 base pairs in length, the cognate site was flanked by slightly different

* To whom correspondence shuold be addressed: Phone: 312-996-0774; e-mail abenight@uic.edu.

‡ Present address: Biotechnology Division, National Institute of Standards and Technology, 100 Bureau Dr., MS 8311, Gaithersburg, MD 20899-8311.

10.1021/bi000326k CCC: $19.00
Published on Web 06/09/2000

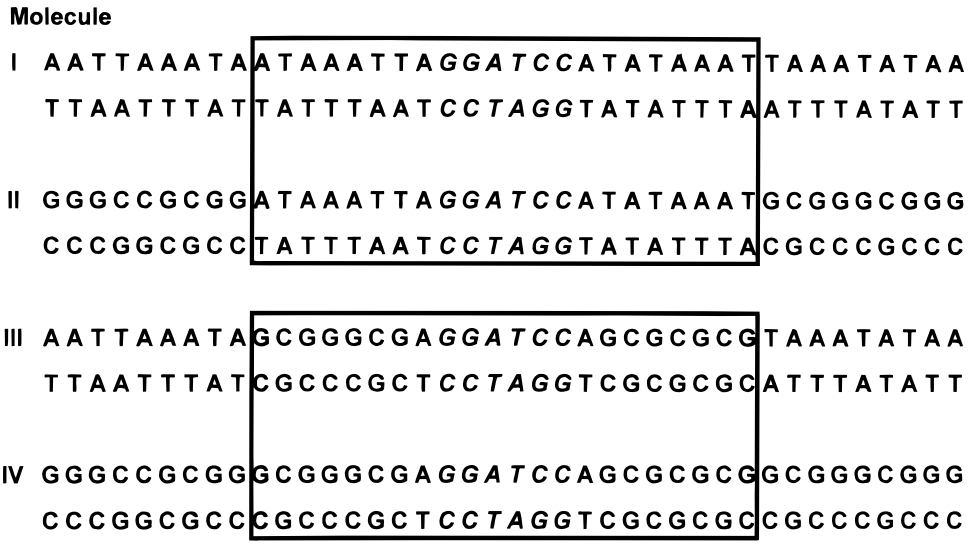

FIGURE 1: Sequences of the 40 base pair DNA Duplexes of this Study. The four DNAs contain two sequence modules A or B, flanked on both sides by 9 base pairs of either 0 or 100% G·C-type base pairs. Each 40 base pair duplex, I−IV, has the same central eight base pair sequence 5-A-*GGATCC*-A-3′, which is the six base pair recognition sequence of *Bam*HI with an A·T base pair on both sides. The central sequence modules are 22 base pairs in length and correspond to the sequences of two 22 base pair duplexes analyzed in a previous study (*6*).

sequence environments having variable thermodynamic stabilities (*1*). Binding-limited cleavage rates were found to vary in a linear manner with sequence-dependent melting of the duplex substrates. Results suggested that site-specific restriction enzyme binding was modulated incrementally by the differential melting stability of sequences flanking the site. However, the flanking sequences in the molecules contained only A·T-type base pairs.

Recently we reported studies of flanking sequences with greater sequence variability on equilibrium binding of *Bam*HI endonuclease (in the absence of Mg^{2+}) (*6*). For these studies a gel-shift assay was developed and employed to examine *Bam*HI binding to six 22 base pair oligoduplex DNAs. The center of each of these duplexes contained the common eight base pair central sequence *5′*-A*GGATCC*A-3′ that includes the *Bam*HI cognate site (italic type). Design of the flanking sequences was such that they varied in the fraction of G·C-type base pairs in different contexts. Results for the six 22 base pair duplexes were analogous to those obtained earlier, and an inverse relationship between flanking sequence stability and *Bam*HI binding was found. That is, sequences with higher free energies of melting (less stable as demonstrated by optical and DSC measurements) exhibited much stronger binding by the enzyme. For the 22 base pair sequences examined with purely AT, purely GC, or AT/GC mixtures in flanking sequences, the inverse relationship between their binding and melting free energies was not strictly linear as found earlier (*1*) for flanking sequences, of purely AT composition in different contexts.

In this paper we report results of investigations of the equilibrium binding (in the absence of Mg^{2+}) of *Bam*HI to its cognate site in considerably longer duplex DNA oligomers. The lengths of sequences flanking the cognate site were increased from seven base pairs in the 22 base pair DNAs studied earlier (*6*) to 16 base pairs in the 40 base pair duplexes studied here. The aim of studying the 40 base pair duplexes was to further investigate relationships between flanking sequence stability and equilibrium binding. For 13 short duplex DNA oligomers ranging in length from 12 to 22 base pairs, an inverse relationship between stability and binding was demonstrated (*1*, *6*). The focus of the present study was to investigate effects of increased flanking sequence lengths, of different composition and context, on equilibrium binding of *Bam*HI to the same site as in the DNAs of our earlier study (*6*). In addition, the four 40 base pair DNAs were thoroughly characterized by circular dichroism spectroscopy and optical and DSC melting analysis. Results for these longer flanking sequences fail to reveal any obvious relationship between equilibrium binding and melting stability.

MATERIALS AND METHODS

DNA Molecules. The eight single strands that comprise the four duplexes shown in Figure 1 were purchased from Oligos Etc. (Wilsonville, OR) and received after being purified by denaturing gel electrophoresis. Prior to shipment, the purity of each strand was confirmed by gel electrophoretic analysis performed by the supplier. On the basis of concentration estimates determined from nearest-neighbor extinction coefficients (*8*), strands were mixed in a 1:1 ratio and purified on nondenaturing gels to remove excess single strands.

Sequences of the four duplexes I−IV that were prepared and studied are depicted in Figure 1. Each 40 base pair duplex contained the six base pair *Bam*HI cognate recognition site, 5′-*GGATCC*- 3′, with an A·T base pair on both sides. This eight base pair site was conserved in the center of all four duplexes. These DNAs are an extension of two of the 22 base pair duplexes studied previously (*6*). The 22 base pair DNAs (duplexes IA and IIIB from that work) form the two sequence modules A and B of the 40 base pair DNAs. The sequences of these 22 base pair modules are highlighted by the boxes in Figure 1. The module A sequence is 18.2% GC, while module B is 81.8% GC.

In DNAs I and II, module A is flanked by two nine base pair sequences having either 0 or 100% GC content. For DNAs III and IV, module B is flanked by the same nine base pair sequences. The total fraction of G·C type base pairs

(f_{GC}) in duplexes I, II, III, and IV is 0.10, 0.55, 0.45, and 0.90, respectively. All experiments were carried out in buffered sodium ion solutions containing 105 mM NaCl, 10 mM sodium cacodylate, and 0.1 mM EDTA, pH = 7.0.

Optical Melting Studies. Optical melting experiments were performed on a Hewlett-Packard 8452 diode-array single beam spectrophotometer. Sample cell temperature was controlled by a Hewlett-Packard 89090A peltier temperature control module. Sample temperature was measured directly with an external temperature probe immersed directly in the solution. Quartz cuvettes with either 1.0 or 0.1 cm path lengths were employed.

To prepare samples for melting curves they were heated to 97 °C for 5 min, then slowly cooled to room temperature, and incubated for at least 4 h. DNA solutions were then filtered through a 0.45 μm nylon filter directly into clean dry cuvettes. Solutions were degassed by bubbling with a fine stream of helium through a submerged micropipet tip for at least 30 min. To prevent evaporation, a drop of pure mineral oil was placed on top of the DNA solution. The cuvette was capped with the external probe and tightly sealed with Teflon tape. After cuvettes were placed in the sample holder, samples were allowed to sit at the starting temperature for at least 30 min.

Absorbance and temperature data points were collected and stored on a PC interfaced to the spectrophotometer by use of resident software provided by Hewlett-Packard. Heating rates were ~35 °C/h with a temperature increment of 0.1 °C. After each temperature was reached, there was a hold time of 6 s. The raw absorbance at 268 nm, the temperature, and the time each data point was collected were used in the analysis. A typical melting experiment consisted of three forward and three reverse scans (heating−cooling curves).

Data points were evenly spaced at 0.1 °C increments by interpolation. A buffer baseline, the absorbance scan of the buffer alone as a function of temperature, was subtracted from the raw absorbance versus temperature curves. The buffer-corrected curves were converted to relative absorbance versus temperature curves by dividing every absorbance value by the first value, subtracting one from each relative value, and plotting the resulting points versus temperature.

Relative absorbance versus temperature curves were normalized to upper and lower linear baselines to determine the fraction of broken base pairs, $\theta_B(T)$, as a function of temperature (*9*). Optical melting curves for the four DNA duplexes were measured as a function of DNA concentration over the approximately 80-fold range from 2.6×10^{-7} to 2.1×10^{-5} M. For these bimolecular melting reactions values of T_m were determined as the temperature at which $\theta = 0.5$, not at $(d\theta_B/dT)_{max}$ as was done for intramolecular melting transitions (*9*). Thermodynamic parameters were evaluated from linear fits of plots of $1/T_m$ versus $ln\ (C_T/4)$ to (*10*)

$$T_m^{-1} = R/\Delta H[ln\ (C_T/4)] + \Delta S/\Delta H \quad (1)$$

Melting Analysis by Differential Scanning Calorimetry. Data collected for the four oligoduplexes of this study were obtained with a Nano-Differential Scanning Calorimeter II Model 6100 (Calorimetry Sciences Corporation, Provo, UT). Sample volumes were approximately 0.6 mL. Prior to temperature scanning, samples were degassed for at least 30 min. Degassing was performed either by bubbling a fine stream of helium through the sample or by placing the sample under house vacuum in a closed vessel.

Sample and reference solutions were loaded into the sample and reference cells with a blunt-tip syringe (Hamilton Gastight 81330). Extreme care was taken in loading solutions to minimize introduction of air bubbles. After both cells were filled, they were capped and a slight external pressure (~3 atm) was applied to prevent evaporation. Raw data were collected as microwatts versus temperature, transferred to a PC, and saved for further analysis.

The accessible temperature range in DSC melting experiments was 5−125 °C. The specific temperature range for each molecule was estimated from optical melting experiments (40−105 °C). DSC melting curves were collected at a scan rate of 60 °C/h. Data points were acquired every 6.25 s. A total of eight scans (four forward and four reverse) were collected for each duplex. Control experiments (not shown) indicated that the ΔH_{cal} values did not vary when DSC data was collected at heating rates of 45 or 90 °C/h. After experiments, samples were removed and the cells were cleaned with 1 L of dilute detergent and flushed with 1 L of Nanopure water.

Analysis of DSC data was performed with resident software packages supplied with the calorimeter. First, the average DNA versus buffer and buffer versus buffer scans were plotted on the same plot. It should be noted that, in the pre- and posttransition regions for the DNA/buffer and buffer/buffer scans, the absolute heat capacity of the DNA solution was lower than the heat capacity of the buffer. This occurred at temperatures outside the region where the DNA exhibited an apparent excess heat capacity due to the enthalpy of melting. After examination of the scans, the buffer versus buffer scan was subtracted from the DNA versus buffer scan, which resulted in baseline-corrected ΔC_p^{ex} versus temperature curves. These baseline-corrected ΔC_p^{ex} versus temperature curves were then normalized for total DNA strand concentration. A second-order progressive baseline was fit to linear sections in the lowest and highest temperature regions of these buffer baseline-corrected curves. Slopes of the progressive baseline changed only slightly for different sequences, depending on the linearity of the low- and high-temperature regions of DSC curves. After subtraction of the fitted baseline, an excess heat capacity curve ΔC_p^{ex} versus temperature [where $\Delta\Delta C_p^{ex} = \Delta C_p^{ex}$ (pre) $- \Delta C_p^{ex}$(post) = 0] remained. This assumption is consistent with the observation that the posttransition ΔC_p^{ex}(post) reading was only slightly higher (5−7%) than the pretransition ΔC_p^{ex}(pre) value. Under these conditions it was very difficult to detect an appreciable $\Delta\Delta C_p^{ex}$ for the melting transitions and thus it was assumed that $\Delta\Delta C_p^{ex} = 0$ for the melting transitions. Resulting thermodynamic parameters ΔH_{cal} and ΔS_{cal} were determined from the integrated areas of excess heat capacity versus temperature curves.

The DSC T_m was determined as the temperature at the peak height maximum on each baseline-corrected ΔC_p^{ex} versus temperature curve. The integrated area under the concentration corrected curve provided the reported DSC transition enthalpy,

$$\int \Delta C_p^{ex}\, dT = \Delta H_{cal} \quad (2)$$

The DSC transition entropy values were determined from the integrated area under curves of $\Delta C_p^{ex}/T$ versus temperature:

$$\int \Delta C_p^{ex}/T \, dT = \Delta S_{cal} \quad (3)$$

The above analysis assumes that the measured transition enthalpy, ΔH_{cal}, and entropy, ΔS_{cal}, are temperature-independent ($\Delta\Delta C_p^{ex} = 0$).

Circular Dichroism Measurements. Circular dichroism spectra were collected on a Jasco J-600 spectropolarimeter at temperatures controlled by a thermostated sample holder regulated by a circulating water bath. DNA solutions were placed in a 0.1 cm path length cell and allowed to equilibrate for at least 15 min before collection of the CD spectrum from 320 to 200 nm at a rate of 10 nm/min. Points were acquired every 0.4 nm with a 2 s time constant and sensitivity of 10−20 mdeg. Plots of millidegrees versus wavelength were derived from the data collected for each sample. At least five scans were acquired and averaged. To account for sample concentration, CD spectra in units of millidegrees were converted to $\Delta\epsilon_{L-R}$. For this conversion the relationship $\Delta A_{L-R} = 4\pi\theta$(degrees)/(180 *ln* 10) or (θ millidegrees/32 982) was used. ΔA_{L-R} was divided by the duplex concentration, resulting in $\Delta\epsilon_{L-R}$ (*11*). This correction for sample concentration gives more common units of $\Delta\epsilon_{L-R}$. After raw data were converted to $\Delta\epsilon_{L-R}$ versus wavelength, relative comparisons of the spectra were made.

To quantify characteristics of the collected CD spectra, singular value decomposition (SVD) was employed. A number of variations on this technique, applied to analyze CD spectral data from DNA and proteins, have been reported (*12*). In our analysis the experimental CD spectra were treated as a set of vectors, and from those data alone, without imposing additional constraints, SVD was used to find an orthogonal set of basis vectors or subspectra that could be linearly combined to construct the original set of spectra. Application of this analysis provided a means for determining, in a least-squares sense, the minimum number of basis spectra with significant weight that could be linearly combined to fit each measured CD spectrum within certain limits. The basis spectra and their linear coefficients were determined from SVD analysis of sets of the collected CD spectra essentially as done previously (*13*).

Equilibrium Binding Studies. The [gel] electrophoretic mobility shift assay (EMSA) was employed to analyze binding reactions of the four 40 base pair DNAs and *Bam*HI restriction endonuclease (*14*−*17*). The *Bam*HI used in these studies was the generous gift of Dr. Ira Schildkraut (New England Biolabs) and was isolated as described (*18*). Concentrated protein (4 mg/mL dimer, 49 kDa, determined by the Bradford assay) was stored in 500 mM potassium chloride, 20 mM potassium phosphate, 1 mM dithiothreitol, and 10% glycerol, pH 6.9. Concentrated protein stock solutions were divided into 15 μL aliquots and stored at −70 °C. Protein dilution buffer was prepared fresh prior to each experiment and contained 10 mM Tris-HCl, pH 7.0 at 25 °C, 1 mM dithiothreitol, 0.1 mM EDTA, and 10% glycerol. Buffer was kept on ice prior to use. *Bam*HI binding buffer consisted of 105 mM sodium chloride, 10 mM sodium cacodylate, and 0.1 mM EDTA, pH 7.0.

Analytical nondenaturing gels were used in the EMSA. To prepare gels for assays, acrylamide solutions were degassed for at least 30 min before polymerization. After being poured, gels were equilibrated at 20 °C in buffer tanks for at least 3 h. Prior to loading samples, gels were preelectrophoresed until a constant current was achieved. Enzymatic radiolabeling procedures were employed as previously described (*6*, *19*). In summary, strands for each of the four duplexes were synthesized without a 5′ phosphate group. This allowed for radiolabeling with ^{32}P. A mixture of the desired amount (2 pmol) of pure single-strand oligonucleotides containing 5′-OH termini, [γ-^{32}P]ATP (6000 Ci/mmol, 1.74 μM ATP concentration, New England Nuclear) in 8-fold excess of the single-strand concentration, and 10−15 units of T4 polynucleotide kinase (New England Biolabs) was prepared in the supplier's 1× kinase buffer. Typical reaction volumes were 15 μL. The mixture was incubated at 37 °C for at least 45 min and not longer than 1 h. To stop the reaction, 2.5 μL of 0.1 M EDTA and 2.5 μL of 8 M urea were added and the solution was heated to 97 °C for 10 min. Labeled strands were purified electrophoretically on 20% analytical denaturing polyacrylamide gels essentially as described (*6*). Binding reactions were carried out in autoclaved, 0.5 mL microcentrifuge tubes. Reaction mixtures were prepared by combining the appropriate volumes of 5× *Bam*HI binding buffer and radiolabeled duplex DNA and incubating at 20 °C for at least 30 min. *Bam*HI enzyme was serially diluted to the appropriate concentrations (10^{-12} to 10^{-6} M) for the assay and kept on ice until used. Aliquots of enzyme were carefully added to the buffer−DNA mixtures and allowed to equilibrate at 20 °C for at least 1 h (total volume 15 μL) in a PTC-100 programmable thermal controller (M. J. Research, Watertown, MA). DNA−protein reaction mixtures were not vortexed or centrifuged. Prior to loading of the gel, 2 μL of tracking dye was placed on the inside of the microcentrifuge tube but not directly in the reaction mixture. A loading voltage of 280 V was applied to the gel and each sample was carefully loaded. The tracking dye was gently tapped into the reaction solution just prior to loading the sample. After samples were loaded, the loading voltage was maintained until the tracking dye migrated 0.5−1.0 cm into the gel. Voltage was then reduced to 180 V and electrophoresis continued until the tracking dye migrated 2−5 cm from the bottom of the gel. Gel temperature was maintained by a thermostated cooling bath at 20 °C circulating through the buffer tank.

Gels were exposed to Fuji RX X-ray film for the time required to obtain a detectable signal intensity yet remain in the linear response range of the film. Typically 1−2 days were required to obtain good visualization of the gels with tungsten-coated intensifying screens (Fisher Scientific). Exposures were done at −70 °C. Film was developed with standard photo fixer and developer solutions from Sigma (*19*).

Autoradiograms were scanned on a flatbed scanner (Paper Port 3000) interfaced to a PC. Analysis of autoradiograms was performed with NIH 1.52 image analysis software. Gel intensity patterns were smoothed and background-corrected. Lanes were outlined and plotted. Gel bands were quantified

[1] Abbreviations: CD, circular dichroism; DSC, differential scanning calorimetry; f_{GC}, fraction of G·C-type base pairs; SVD, singular value decomposition; EMSA, electrophoretic mobility shift assay.

in pixels. The numbers of pixels corresponding to free DNA and DNA−protein complex(es) was tabulated at each *Bam*HI concentration. The fractions of each DNA−protein complex were determined by dividing the number of pixels in a complex by the total number of pixels in the entire lane (*6*).

Analysis of Binding Isotherms. The model used to analyze binding isotherms was similar to that reported to evaluate binding constants for cooperative site-specific protein−DNA interactions (*20*). Consider the binding equilibrium

$$\mathrm{D} + a\mathrm{E} \overset{K_{\mathrm{B1}}}{\rightleftharpoons} \mathrm{DE}_a + b\mathrm{E} \overset{K_{\mathrm{B2}}}{\rightleftharpoons} \mathrm{DE}_b \tag{3}$$

In this two-step binding scheme D is the free duplex DNA, E is *Bam*HI dimer, *a* and *b* are stoichiometric coefficients that represent the numbers of dimers bound, and K_{B1} and K_{B2} are the primary and secondary binding constants, respectively. In the first step, free duplex, D, is bound by *a* enzyme dimers, E, to form the primary complex, DE_a. Presumably this step corresponds to binding primarily at the cognate site and is considered to be sequence-specific. The primary bound species can then be subsequently bound by an additional *b* enzyme dimers to form the secondary complex, DE_b. Binding in this mode is assumed to be less sequence-specific. The equilibrium association constants for each step of the binding reaction, K_{B1} and K_{B2}, are given by

$$K_{\mathrm{B1}} = \frac{\mathrm{DE}_a}{\mathrm{D[E]}^a} \tag{4a}$$

$$K_{\mathrm{B2}} = \frac{\mathrm{DE}_b}{\mathrm{DE}_a\mathrm{[E]}^b} \tag{4b}$$

Assigning the free duplex as the reference state, the partition function, *Q*, for the binding reaction is given by

$$Q = 1 + (K_{\mathrm{B1}}E)^a + (K_{\mathrm{B1}}K_{\mathrm{B2}}E)^{(a+b)} \tag{5}$$

With this definition, the fraction of duplex bound in the primary mode (first shifted band) is given by

$$F_{\mathrm{B1}} = \frac{(K_{\mathrm{B1}}E)^a}{1 + (K_{\mathrm{B1}}\mathrm{E})^a + (K_{\mathrm{B1}}K_{\mathrm{B2}}\mathrm{E})^{(a+b)}} \tag{6}$$

Plots of the experimentally determined values of F_{B1} versus log of the *Bam*HI concentration were fit with eq 6 by varying the primary binding constant, K_{B1}, as an adjustable fitting parameter. The overall association constant, $K_{\mathrm{B1}}K_{\mathrm{B2}}$, was also allowed to vary. Values of the coefficients *a* and *b* were set equal in the fits. Fitting was performed by nonlinear least-squares and provided semiquantitative estimates on the primary mode binding constant, K_{B1}, and its standard deviation for each of the four DNA duplexes. Reasonable fits to the primary mode binding data were obtained with this model and thereby provided a means of comparing effects of different flanking sequences on *Bam*HI equilibrium binding.

RESULTS

Optical Melting Experiments. Ultraviolet absorbance (λ = 268 nm) versus temperature melting curves were collected as a function of total strand concentration, C_{T}, from approximately 2.6×10^{-7} to 2.1×10^{-5} M for the DNAs of Figure 1. The averages of three such curves obtained for each duplex at a strand concentration of 6.0 μM are shown in Figure 2a, where θ_{B} (fraction of broken base pairs) is plotted versus temperature. Plots of θ_{B} versus temperature were comparable in shape at all strand concentrations examined. On these curves there was no evidence of biphasic or otherwise abnormal behavior. The hyperchromic shifts observed for the four duplexes were as follows: DNA I (26.0%), DNA II (24.2%), DNA III (21.3%), and DNA IV (24.4%). The T_{m}'s determined at 6.0 μM are listed in Table 1. The comparisons in Table 1 and Figure 2a show (not surprisingly), at identical strand concentrations, differences in T_{m}'s of the DNAs generally increase with increasing differences in the percent GC of their sequences. At 6.0 μM, T_{m}'s of the duplexes varied over a 33.9 °C range. Figure 2a depicts the data for DNA I (●), DNA II (○), DNA III (■), and DNA IV (□). The sequences of these DNAs have 10% (I), 55% (II), 45% (III), and 90% (IV) G·C-type base pairs. Sequences II and III have similar total percentages of GC (55% and 45%, respectively), but differ in their core sequence

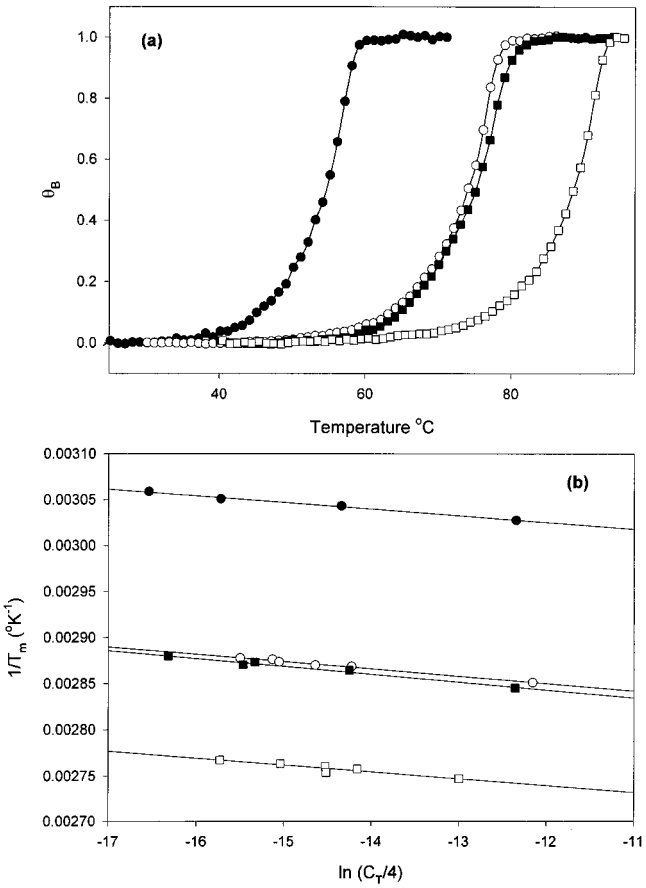

FIGURE 2: Results of optical melting experiments for the four 40 base pair duplexes. (a) Normalized melting curves of plots of the fraction of broken base pairs, θ_{B}, versus temperature for the four DNA duplexes of Figure 1. These curves were collected at an identical strand concentration of 1×10^{-6} M and are the averages from at least three independent melting experiments. Symbols are as follows: DNA I (●), DNA II (○), DNA III (■), and DNA IV (□). Values of the optical transition temperature, T_{m}, derived from these curves are summarized in Table 1. (b) van't Hoff plots for the 40 base pair DNA duplexes. Plots of ln (C_{T}/4) versus $1/T_{\mathrm{m}}$ were constructed from the concentration dependence of T_{m} values, determined in optical melting experiments. Plots shown are for DNA I (●), DNA II (○), DNA III (■), and DNA IV (□). The linear fits that are shown provided estimates of the van't Hoff enthalpy, ΔH_{vH}, and entropy, ΔS_{vH}, as summarized in Table 2.

Table 1: Thermodynamic Parameters for Four 40-Base-Pair Duplexes Derived from Optical Melting Experiments

DNA	f_{GC}	% hyp	T_m (°C) (6.0 μM)	ΔH (kcal/mol)	ΔS [cal/(mol K)]	$\Delta G_{293.15}$ (kcal/mol)
I	0.10	26.0	54.7 ± 0.2	−273.4 ± 16.1	−803.3 ± 49.1	−37.9 ± 1.7
II	0.55	24.2	74.2 ± 0.4	−246.6 ± 12.4	−678.9 ± 35.6	−47.6 ± 2.0
III	0.45	21.3	75.2 ± 0.4	−230.5 ± 20.0	−631.5 ± 57.3	−45.4 ± 3.2
IV	0.90	24.4	88.6 ± 0.3	−262.7 ± 56.7	−695.8 ± 156.3	−58.8 ± 10.9

modules. Interestingly, the T_m value for DNA III (with module B) containing a 22 base pair core (81.8% GC) flanked by nine A·T base pairs is higher than DNA II with an AT-rich 22 base pair core (18.2% GC) flanked by nine G·C base pairs. The T_m for DNA III is higher than that for DNA II by only 1 °C (at $C_T = 6.0$ μM), a clear indication of different sequence context effects on T_m. Apparently, not only the sequence content (percent GC) but also the order, or context, of the base pairs determines the stability of these DNAs. According to the T_m values the order of stability of the four duplexes is IV > III > II > I.

van't Hoff plots of ln (C_T/4) versus $1/T_m$ were constructed for the four duplexes from the concentration dependence of their T_m values. These plots are shown in Figure 2b for DNA I (●), DNA II (○), DNA III (■), and DNA IV (□) and demonstrate the concentration-dependent melting properties of the DNAs as a function of sequence. Linear fits to the data provided estimates of the van't Hoff enthalpy, ΔH_{vH}, and entropy, ΔS_{vH} . Application of this method is based on the underlying assumption that the melting transition occurs in a two-state manner. In all likelihood this assumption is only marginally valid for DNAs of this length. Nonetheless the graphically evaluated two-state thermodynamic parameters are summarized in Table 1. As shown there, ΔH_{vH} values range from −230.5 kcal/mol for DNA III to −273.4 kcal/mol for DNA I. Similarly, the van't Hoff entropy values, ΔS_{vH}, range from −631.5 eu for DNA III to −803.3 eu for DNA I. According to their melting free energies, the order of stability of the four duplexes is IV (−58.8 kcal/mol) > II (−47.6 kcal/mol) ≥ III (−45.4 kcal/mol) > I (−37.9 kcal/mol). Because of the questionable reliability of the transition thermodynamic parameters evaluated from optical melting curves due to shortcomings of the two-state assumption, a model-independent analysis was required. For this purpose, DSC melting experiments were performed.

DSC Melting Experiments. Results from DSC melting experiments for DNA I (●), DNA II (○), DNA III (■), and DNA IV (□) are shown in Figure 3, where excess heat capacity, ΔC_p^{ex}, is plotted versus temperature. DSC scans were obtained at approximately the same total strand concentration (∼2.5×10^{-4} M). The curves in Figure 3 show clearly resolved melting transitions for the four DNA molecules. The DSC curves for the module A duplexes (DNA I and DNA II) although separated by about 20 °C, have similar shapes, higher peaks, and sharper transitions than the melting transitions of the module B DNAs (DNA III and DNA IV). Apparently, the AT-rich 22 base pair core of the duplexes having module A results in higher peaked, more cooperative melting transitions. In contrast, the melting transitions for DNAs III and IV having the GC-rich module B core, are broader and less cooperative. This is especially apparent for DNA III (■), whose DSC curve displays a significant premelting transition from approximately 60 to 78 °C. This broad pretransition probably corresponds to

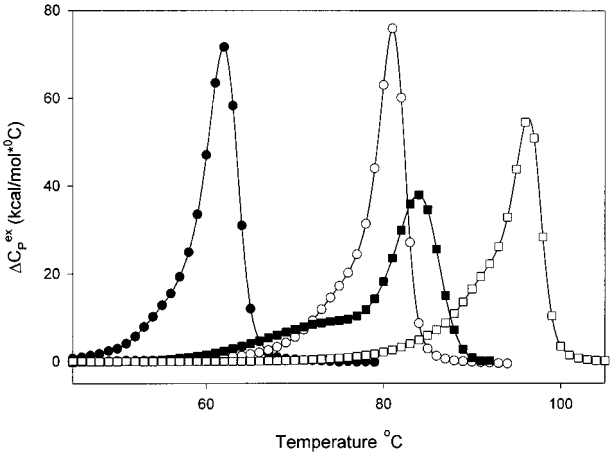

FIGURE 3: Results of DSC melting. DSC melting curves, plots of the excess heat capacity, ΔC_p^{ex}, versus temperature for DNA I (●), DNA II (○), DNA III (■), and DNA IV (□) are shown. DSC experiments were preformed for each DNA duplex at similar strand concentrations (∼$2.5 \times 10^{-4} \pm 0.6 \times 10^{-4}$ M).

Table 2: Thermodynamic Parameters for Four 40 Base Pair Duplexes Derived from DSC Experiments

DNA	ΔH (kcal/mol)	ΔS [cal/(mol K)]	$\Delta G_{293.15}$ (kcal/mol)
I	−442.0 ± 4.8	−1330.0 ± 14.7	−52.2 ± 0.5
II	−429.2 ± 9.6	−1221.9 ± 27.5	−71.0 ± 0.6
III	−390.5 ± 13.3	−1094.6 ± 33.8	−65.5 ± 0.5
IV	−389.1 ± 17.2	−1062.7 ± 44.7	−77.5 ± 0.7

melting of the A·T base pairs on the ends prior to melting of the GC-rich duplex core of DNA III. The DSC T_m values were determined from the temperature at the peak height maximum on these curves. Values of the DSC transition enthalpy and entropy, ΔH_{cal} and ΔS_{cal}, obtained from these curves are summarized in Table 2. As shown there, DNA I exhibited the highest ΔH_{cal} at −442.0 kcal/mol while DNA IV had the lowest at $\Delta H_{cal} = -389.1$ kcal/mol. The entropy values range from −1330 cal/K·mol for DNA I to −1063 cal/K·mol for DNA IV. Both the enthalpy and entropies are smaller for the module B duplexes. The free energies, $\Delta G_{293.15}$, of the duplexes increased with decreasing f_{GC} from −77.5 for DNA IV to −52.2 for DNA I.

Recently, several authors have reported that there may be a significant difference in the excess heat capacities of the duplex and single strand states, $\Delta\Delta C_p^{ex} \neq 0$. It was also reported that measurements of this difference by DSC can be difficult (*25*, *26*). We also experienced difficulty in consistently detecting meaningful values of $\Delta\Delta C_p^{ex}$ in our DSC melting experiments. As stated earlier, the post transition ΔC_p^{ex} readings were only slightly higher (<7%) than the pretransition ΔC_p^{ex} values, but this varied for different duplexes. These relatively small differences combined with uncertainties associated with choices of pre- and posttransition baseline regions on melting curves, and the extreme

Table 3: Comparisons of Thermodynamic Parameters Derived from Optical and DSC Melting Experiments for Four 40 Base Pair Duplexes

DNA	$\Delta\Delta H$ (kcal/mol)	$\Delta\Delta S$ [cal/(mol K)]	$\Delta\Delta G_{293.15}$ (kcal/mol)
I	-168.6 ± 16.8	-526.7 ± 51.3	-14.3 ± 1.8
II	-182.6 ± 15.6	-543.0 ± 45.0	-23.4 ± 2.1
III	-160.0 ± 24.0	-463.1 ± 66.5	-20.1 ± 10.5
IV	-126.4 ± 59.3	-366.9 ± 162.6	-19.0 ± 10.9

sensitivity of the final results to these choices, made it difficult to accurately discern an appreciable $\Delta\Delta C_p^{ex}$ on DSC melting curves. Thus, under the conditions employed an appreciable difference between ΔC_p^{ex} for the duplex and single-strand states could not be determined and for these reasons the most consistent results, those obtained with the assumption that, $\Delta\Delta C_p = 0$, are reported. If $\Delta\Delta C_p^{ex} \neq 0$ the evaluated transition parameters, ΔH_{cal} and ΔS_{cal}, are temperature-dependent. As discussed below, this could significantly affect values of the free energy calculated at temperatures (i.e., at 20 °C) far below the melting temperature.

The DSC and optical thermodynamic parameters are compared in Table 3. Significant differences are seen that probably result from substantial deviations from two-state melting behavior for the 40 base pair duplex DNAs. This might have been anticipated from the DSC melting curve for DNA III, which displays pretransition melting, a clear sign of non-two-state melting behavior. Deviations from two-state melting are confirmed by the large differences between the optical and DSC parameters found for all four DNAs as shown in Table 3. Because the model-independent DSC values are deemed more reliable, they are used in the following comparisons. The DSC T_m's for the four duplexes agreed to within 1 °C with the T_m's linearly extrapolated from the concentration-dependent optical melting data to the higher concentrations where DSC melting experiments were conducted. Although there are clear quantitative differences between the optical and DSC melting parameters, according to their free energy values, differences in stability for the DNAs determined by either technique are the same.

It is interesting to note for deoxyoligoduplexes with different sequence contexts (DNAs II and III) that the T_m measured in optical melting experiments for DNA III (75.2 °C) is greater than that observed for DNA II (74.2 °C). This is also the case for the T_m values determined by DSC where the T_m of DNA III (84.4 °C) is greater than that for DNA II (80.9 °C) (data not tabulated). This ordering of the T_m values does not strictly follow the f_{GC} of the duplexes, but the relative orders of the free energies follow f_{GC}, at -65.5 kcal/mol for DNA III and -71.0 kcal/mol for DNA II.

Circular Dichroism Spectra. The CD spectra of DNAs, I (●), II (○), III (■), and IV (□) collected over the wavelength range from 200 to 320 nm are shown in Figure 4a. The magnitudes of the negative band around 247 nm and positive band at 210 nm both decrease with increasing f_{GC}.

To obtain a more quantitative assessment of the distinguishable spectral components comprising the spectra in Figure 4a, and their relative changes with sequence, the collected CD spectra were subjected to factor analysis by SVD (*13*, *27*, *28*). Analysis by this procedure supplies the minimum number of subspectra and their relative weight coefficients that can be linearly combined to produce the measured spectra within experimental confidence limits. Factor analysis of the CD spectra in Figure 4a revealed that the original CD spectra for all four oligoduplexes could be reconstructed from linear combinations of one major and two minor subspectra. These subspectra are shown in Figure 4b. Their relative weights were 0.79 (●), 0.17 (○), and 0.03 (▼), which represent the principal spectroscopically distinguishable features of the DNAs. Values of the coefficients of the major subspectra for the four DNA molecules were used as a gauge of incremental sequence-dependent variations of the

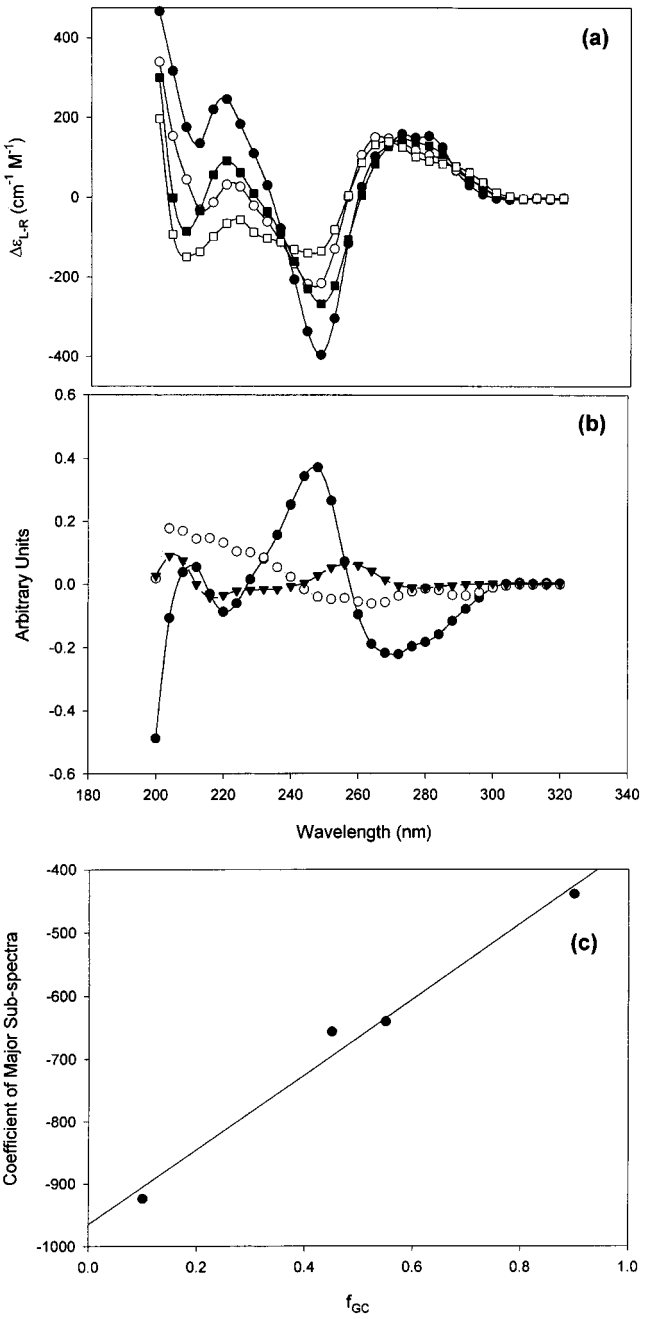

FIGURE 4: Results of circular dichroism measurements. (a) The CD spectra for the four DNAs, collected over the wavelength range from 200 to 320 nm. The spectra shown are for DNA I (●), DNA II (○), DNA III (■), and DNA IV (□). (b) Subspectra derived from SVD analysis of the four spectra in panel a. Three subspectra with relative weights of 0.79 (●), 0.17 (○) and 0.03 (▼) are shown. (c) Coefficient of the major subspectrum versus fraction of G·C base pairs, f_{GC}, in each duplex sequence. Linear changes in this coefficient occur with increasing GC percentage.

Table 4: Comparisons of Predicted and Calorimetric Thermodynamic Parameters

DNA	ΔH_{pred} (kcal/mol)	ΔH_{cal} (kcal/mol)	ΔS_{pred} [cal/(mol K)]	ΔS_{cal} [cal/(mol K)]	$\Delta G_{pred\ 293.15}$ (kcal/mol)	$\Delta G_{cal\ 293.15}$ (kcal/mol)
I	−343.7	−442.0	−994.0	−1330.0	−52.4	−52.2
II	−361.0	−429.2	−994.0	−1221.9	−69.6	−71.0
III	−358.0	−390.5	−994.0	−1094.6	−66.6	−65.5
IV	−376.0	−389.1	−994.0	−1062.7	−84.6	−77.5

DNA	$\Delta\Delta H$ (kcal/mol)	$\Delta\Delta S$ [cal/(mol K)]	$\Delta\Delta G_{int\ 293.15}$ (kcal/mol)
I	−98.3	−336.0	0.20
II	−68.2	−227.9	−1.4
III	−32.5	−100.6	1.1
IV	−13.1	−68.7	7.1

[a] Upper panel: Thermodynamic parameters ΔH, ΔS, and $\Delta G_{293.15}$ were evaluated from DSC melting experiments and predicted from the nearest-neighbor parameters of Owczary et al. (*40*). Lower panel: $\Delta\Delta$ (cal − pred) values are tabulated. $\Delta G_{int\ 293.15} = \Delta G_{cal} - \Delta G_{pred}$.

CD spectra and are plotted versus f_{GC} in Figure 4c. Coefficients of the major subspectrum, which has a strong peak at 247 nm, follow the opposite trend as the depth of the major trough seen at 247 nm on the actual CD spectra and increase with increasing f_{GC}.

Duplex Initiation Free Energies for the 40 Base Pair DNAs. The initiation free energy is defined as $\Delta G_{int} = \Delta G_{cal} - \Delta G_{pred}$, where ΔG_{cal} is the DSC-obtained free energy and ΔG_{pred} is the predicted free energy. The values of ΔG_{pred} determined at 20 °C, $\Delta G_{pred\ 293.15}$, for DNAs I−IV are given in Table 4 and compared directly with the $\Delta G_{cal\ 293.15}$ values calculated at 20 °C from the experimentally measured values of ΔH_{cal} and ΔS_{cal}, assuming they are temperature-independent. The $\Delta G_{cal\ 293.15}$ values include contributions from both the favorable sequence dependent (predicted) free energies and the unfavorable energetic cost of duplex initiation. The $\Delta G_{pred\ 293.15}$ values for the four duplex sequences can be readily determined from any of the published sets of nearest-neighbor sequence-dependent stability parameters (*29*−*39*). A number of these sets have been reported and recently reviewed (*40*, *41*). For our calculations we employed the parameter set evaluated from melting studies of DNA dumbbells (*40*). This set seemed the most appropriate because it was determined in the same Na^+ environment (115 mM) in which the four duplexes of the present study were melted. As Table 4 shows, the values of ΔG_{int} are similar for the duplexes containing the AT-rich core of module A. These ΔG_{int} values (+0.2 kcal/mol for DNA I and −1.4 kcal/mol for DNA II) are within 1.5 kcal/mol of ΔG_{cal} and essentially in agreement within the error of DSC experiments. However, for duplexes containing the GC rich module B sequence (DNA III and DNA IV), $\Delta G_{int} = +1.1$ and +7.1 kcal/mol, respectively. These significantly higher values of ΔG_{int} reveal substantial deviations from the predicted stability for the module B duplexes. That is, assuming ΔG_{int} does not depend on sequence, the module A duplexes (DNAs I and II) behave in a predictable manner, while the module B DNAs (DNA III and IV), containing the GC-rich 22 base core, deviate from prediction. For DNA III, where the ends flanking module B are composed of A·T-type base pairs, the deviation from prediction is less than that for DNA IV which contains all G·C base pairs flanking the module. CD spectra and PAGE characterization of the duplexes did not reveal the presence of any anomalous structures or odd properties of DNAs III and IV that would suggest difficulty in predicting their sequence-dependent stabilities. Even though DNA III exhibited a broad shoulder on its melting transition, its predicted free energy was more accurate than that for DNA IV. There are several potential reasons for these findings.

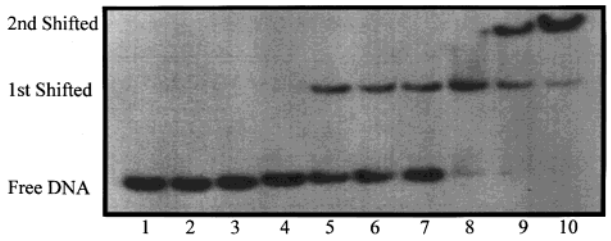

FIGURE 5: Typical result from gel mobility shift assays. The autoradiogram is typical of experimental data collected from mobility shift assays, in which *Bam*HI was titered against a constant concentration of the four duplex substrates in Figure 1. The particular result shown is for DNA II. Reaction components in each of the lanes are as follows: lane 1, DNA II alone at 1.0×10^{-10} M (no enzyme added); lanes 2−10, contained 1.0×10^{-9}, 1.0×10^{-8}, 3.0×10^{-8}, 5.0×10^{-8}, 7.0×10^{-8}, 1.0×10^{-7}, 3.0×10^{-7}, 5.0×10^{-7}, and 7.0×10^{-7} M *Bam*HI dimers, respectively, in addition to DNA duplex at 1.0×10^{-10} M.

As stated above, if the melting process of the molecules involves significant differences between the excess heat capacities of the duplex and single strand states ($\Delta\Delta C_p^{ex} \neq 0$), then extrapolation to 20 °C could lead to significant differences in the calculated free energies. The fact that relative differences in the values of ΔG_{int} are found for different sequences suggests if $\Delta\Delta C_p^{ex}$ is nonzero it is also sequence-dependent. In addition, difficulties of the nearest−neighbor model in accurately predicting the stabilities of DNAs III and IV may be due to the presence of significant long range effects in the duplex sequences of these DNAs that are not considered by the nearest-neighbor model. Further, the different sequences may reside in different structural states with different stabilities that could contribute to departures from nearest-neighbor predictions. In particular, the apparently anomalous melting behavior of DNA IV suggests longer-range interactions and/or alternate structures could exist in long stretches of G·C-type base pairs.

Given the aforementioned possibilities, the significant differences in the values obtained for ΔG_{int} suggest the sequence dependence of ΔG_{int} may not arise solely from sequence-dependent nucleation at the duplex ends. Instead, identity of the core module of the duplex is apparently the variable sequence feature that correlates with a significant value of ΔG_{int}. This finding favors the argument that factors other than end effects are responsible for the observed sequence-dependent deviations from the melting behavior predicted from the nearest−neighbor model.

BamHI Equilibrium Binding. Equilibrium binding of *Bam*HI restriction endonuclease to the four 40 base pair duplexes was monitored in electrophoretic mobility shift assays. A typical example of the experimental data is shown in Figure 5. The particular autoradiogram shown was acquired from a binding experiment with DNA II. Similar data were collected for the other DNAs (not shown). Lanes for the gel experiment depicted in Figure 5 contained DNA II duplex alone (no enzyme added) in lane 1 and DNA II at the same concentration but in the presence of increasing concentrations of *Bam*HI (lanes 2−10).

As shown in Figure 5, three bands were observed and designated free DNA (unbound duplex), first shifted material, and second shifted material. On some gels a fourth high-mobility band was observed (not seen on the gel in Figure 5) and was attributed to an excess of labeled free single-strand DNA. In most cases, as under the conditions of Figure 1, the free single strand ran off the gel. When desired, this species could be clearly visualized by use of shorter electrophoresis times. The electrophoretic mobility patterns revealed two interrelated modes of equilibrium binding by *Bam*HI. Because of the different order of their onset and for descriptive purposes, these binding modes are designated as the primary and secondary modes. The primary binding mode corresponds to the first shifted band, occurs at lower enzyme concentrations, and presumably reflects sequence-specific binding at the cognate site. The slower mobility, second shifted band was assigned to a secondary binding mode, attributed to binding of more than one *Bam*HI dimer to the 40 base pair duplexes. These assignments are not unprecedented as others have also attributed the presence of multiple retarded mobility bands in gel-shift assays of *Eco*RI and *Bam*HI to a secondary binding mode (*6*, *42*−*45*). Furthermore, preliminary crystallization studies of *Bam*HI/DNA complexes showed that 12 base pairs of duplex are easily able to accommodate a single *Bam*HI dimer (*46*). Therefore, it seems entirely plausible that our 40 base pair duplexes could bind at least two and perhaps even more *Bam*HI dimers in less sequence-specific modes.

As described below, binding curves of F_{B1} versus log [*Bam*HI] for the primary mode of binding were typically bell-shaped and F_{B1} never reached an observable value of 1. As the fraction of secondary mode material increased, the fraction of primary material decreased concomitantly. Thus, the two binding modes overlap. An example of this can be clearly seen in lanes 8−10 of Figure 5. Both the first shifted (primary mode) and second shifted (secondary mode) bands appear in the same lanes (8−10) with varying intensity.

Binding curves for the primary mode (first gel-shifted material) generated from multiple gel-shift assays with the four duplexes are displayed in Figure 6a. On these curves the fraction bound in the primary mode, F_{B1}, is plotted versus log [*Bam*HI]. Symbols for each curve are averages of at least three experiments, and error bars denote the standard deviations for multiple experiments. The curves correspond to DNA I (●), DNA II (○), DNA III (■), and DNA IV (□). The solid curves drawn through the data were obtained by fitting F_{B1} from eq 6 subject to the following constraints. Values of the coefficients *a* and *b* were set equal in the fits and the product $K_{B1}K_{B2}$ was allowed to vary over the range from 1.8×10^6 M^{-1} to 1.0×10^7 M^{-1}. In this way K_{B2} values were constrained to the concentration range over which the fraction of secondary mode binding for the four duplexes undergoes the transition from fully unbound ($F_{B2} = 0.0$) to fully bound ($F_{B2} = 1.0$). This range is defined from plots of the fraction of second shifted material, F_{B2}, versus log [*Bam*HI] as depicted in Figure 6b. This plot shows that all of the second-mode binding for the four duplexes occurs over the relatively narrow enzyme concentration range of $(3.0-9.0) \times 10^{-7}$ M.

For all duplexes, onset of primary binding occurs over the range of enzyme concentrations from about 5.5×10^{-8} to 1.5×10^{-7} M. The order of onset of the primary binding mode can be qualitatively ranked as II > III > I > IV. The values of K_{B1} that were obtained for DNAs I−IV by adjusting K_{B1} in eq 6 to fit the primary isotherms are summarized in Table 5. These values range from 0.6×10^{-7} ($\pm 0.9 \times 10^{-7}$) to 4.2×10^{-7} ($\pm 0.6 \times 10^{-7}$) M. In the fits, values of the stoichiometric coefficients *a* and *b* in eq 6 were set equal to 2, because it was not possible to fit the observed binding behavior within the errors when *a* and *b* were set equal to 1. This reveals a possible shortcoming of the model and/or the limited resolution of the data. Prior to onset of secondary-mode binding, the fractions of primary-mode binding for duplexes having the module A sequences (DNAs I and II)

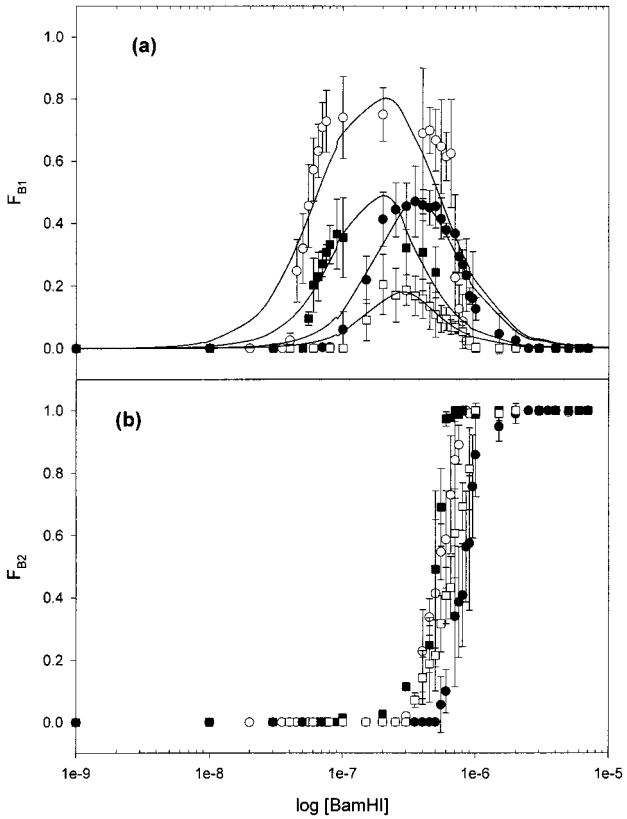

FIGURE 6: Results of binding experiments. (a) Primary binding mode isotherms for the four 40 base pair DNAs. Binding curves for the primary mode (first gel shifted material) are displayed for the four duplexes. The fraction of duplex bound in the primary mode, F_{B1}, is plotted versus log [*Bam*HI]. Symbols for each curve are the averages of at least three experiments, and error bars represent the standard deviation from multiple experiments. The curves correspond to DNA I (●), DNA II (○), DNA III (■), and DNA IV (□). Solid curves drawn through the data were obtained by fitting F_{B1} to eq 6 as described in the text. (b) Secondary binding mode isotherms for the four 40 base pair DNA duplexes. Binding curves for the secondary mode (second gel-shifted material) are displayed for the four duplexes. In these isotherms the fraction bound in the secondary mode, F_{B2}, is plotted versus log [*Bam*HI]. Symbols for each curve are the averages of at least three experiments, and error bars represent the standard deviations from multiple experiments. The curves correspond to DNA I (●), DNA II (○), DNA III (■), DNA IV (□).

Table 5: Parameters for Primary Mode Binding to *Bam*HI

DNA	K_{D1} (M)	$\Delta G_{B1\ 293.15}$
I	$2.7 \times 10^{-7} \pm 0.7 \times 10^{-7}$	-8.8 ± 0.2
II	$0.6 \times 10^{-7} \pm 0.9 \times 10^{-8}$	-9.7 ± 0.2
III	$1.3 \times 10^{-7} \pm 0.6 \times 10^{-7}$	-9.2 ± 0.1
IV	$4.2 \times 10^{-7} \pm 0.6 \times 10^{-7}$	-8.6 ± 0.1

reach higher maximum values (0.75 and 0.45) than the primary mode fractions (0.40 and 0.20) for the module B duplexes. Thus, the module B DNAs (III and IV), with G−C-rich cores, see the onset of the secondary mode of binding before the fraction involved in primary-mode binding reaches 0.2 (in the case of DNA IV). That secondary-mode binding occurs over such a relatively narrow range of *Bam*HI concentrations (3.0×10^{-7} to 9.0×10^{-7} M) suggests that this binding mode occurs in a highly cooperative manner. Qualitatively, the relative peak height maxima on the plots in Figure 6a indicate the relative strengths of the primary and secondary binding modes of *Bam*HI for each of the four sequences. The lower peak heights in Figure 6a for the duplexes having internal G·C base pairs suggests the secondary mode of binding for these DNAs occurs to a relatively greater extent but at the expense of the primary binding mode. The value of 2 for the stoichiometric coefficients *a* and *b* required to fit the binding curves is also indicative of binding cooperativity. For secondary-mode binding this is evident in the binding isotherms. Secondary-mode binding occurs over a narrow range of *Bam*HI concentrations (less than a factor of 10) and displays steep sigmoidal behavior. Perhaps after primary binding occurs the remaining exposed DNA becomes "primed" for subsequent nonspecific binding of more enzyme dimers. This seems reasonable because of the high enzyme concentrations (10^{-7}−10^{-6} M) where secondary binding is observed. Finally, the factor of 2 between the coefficients ($a = 2$, $a + b = 4$) indicates that the secondary binding mode is significantly more cooperative than primary-mode binding.

DISCUSSION

Comparisons with Previously Studied 22 Base Pair Duplex Oligomers. The four duplexes studied here are an extension of two previously studied 22 base pair duplexes (*6*), which are a subset of the six duplexes that were analyzed earlier. In the former studies, binding of *Bam*HI to the six duplexes was found to correlate in an approximately inverse manner with the evaluated thermodynamic stability of the duplex sequences.

In this study we wished to probe the role and potential relationships between binding and stability in longer DNAs. Our experimental design was motivated by the inquiry of whether similar correlations between *Bam*HI binding and flanking sequence stability could also be found in longer duplex DNA environments. The sequences of the two precursor 22 base pair duplexes are shown in boxes in Figure 1. In our previous study these two duplexes exhibited the largest differences in their behaviors probed by melting, CD, and equilibrium binding by *Bam*HI. Thus we chose to expand upon these 22 base pair duplex sequences for investigations of longer flanking sequence context effects. In our previous study strong correlations between stability estimated by optical and calorimetric melting experiments, CD spectra, and equilibrium binding by *Bam*HI restriction endonuclease were obtained. Here, the extended set of four 40 base pair duplexes was subjected to identical characterizations and results were compared with our previous findings for the 22 base pair DNAs.

Stability of the four 40 base pair duplexes follows the expected trend. With increasing GC percentage the stability increases. This was also found for the six 22 base pair duplexes. For both sets of molecules, thermodynamic transition parameters evaluated from optical and DSC melting experiments followed similar trends qualitatively but were not in quantitative agreement. This finding reinforces the assertion that the melting transitions are non-two state and that a model-independent means is essential to accurately determine the thermodynamic stability. The free energies for the six duplexes were predicted from our set of nearest−neighbor parameters (*40*) and the initiation free energy, ΔG_{int}, was evaluated as done previously (*6*). In the case of the 22 base pair duplexes, there was a significant ΔG_{int} for the six duplexes and values of ΔG_{int} ranged from +11.6 to +15.9 kcal/mol. Values for the 22 base pair core modules A and B were +11.6 and +13.3 kcal/mol, respectively. From the set of 40 base pair duplexes, DNAs I and II were predicted to be within the error of the experiments, while DNAs III and IV deviated by +2.2 and +8.2 kcal/mol, respectively. The extension of the duplexes from 22 to 40 base pairs should decrease contributions from end effects and other oligomeric-based properties that may have caused the much greater ΔG_{int} values determined for the 22 base pair duplexes. Most notable is that departures from two-state melting behavior for the longer DNAs are likely to be greater for the longer DNAs, which could affect the values of ΔG_{int}. The DNAs with module A have a 22 base pair AT-rich core, while module B duplexes have a GC-rich core. Values of ΔG_{int} found for GC rich sequences and not for the AT-rich sequences may reveal the presence of interactions in the G−C rich sequences that extend beyond nearest neighbors, suggesting the different sequences probably reside in different structural states having different stabilities. Also, if the melting process of these molecules involves significant differences between the excess heat capacities of the duplex and single-strand states ($\Delta\Delta C_p^{ex} \neq 0$), then extrapolation to 20 °C could lead to significant differences in the calculated values of ΔG_{int}. If so, then the finding of different values of ΔG_{int} for different sequences suggests that nonzero values of $\Delta\Delta C_p^{ex}$ are also sequence-dependent. Comparison of the derived values of ΔG_{int} in Table 4 with the binding free energies in Table 5 does not reveal any obvious correlations.

The CD spectra for the 22 base pair and 40 base pair DNAs displayed identical trends. That is as the total f_{GC} increased for each duplex, the character of the spectra responded in a linear fashion. Analysis of the CD spectra by SVD revealed for both the 22 and 40 base pair DNAs that three subspectra can be linearly combined to produce the measured spectra, and coefficients of the major subspectrum vary linearly with f_{GC}. The observation of three subspectra could reveal subtle conformational changes (slight variations of B-form DNA) and differences in the CD spectra due to AT/GC content of the duplexes, most pronounced in the major subspectrum. Other than these observations, significant correlations between sequence context and the CD spectra did not emerge. The presence of A·T- or G·C-type base pairs in different orders did not cause significant deviations from linearity in plots of f_{GC} versus the major subspectrum coefficient. Thus, observed differences in the CD spectra are likely due more to the different spectroscopic characteristics of A·T versus G·C base pairs than to significant structural differences of the duplex DNAs.

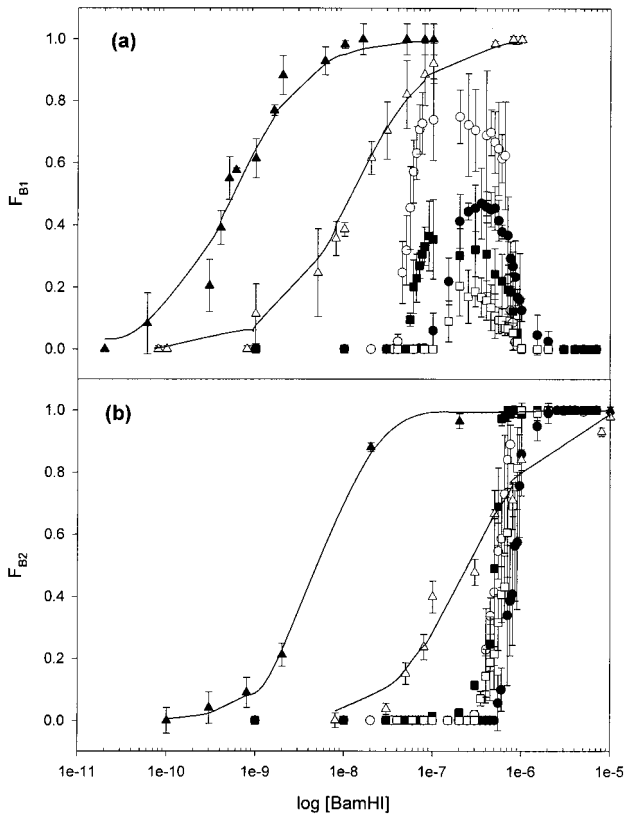

FIGURE 7: Comparison of the binding isotherms for the four 40 base pair duplexes and two 22 base pair DNA duplexes. (a) Binding curves for the primary mode (first gel shifted material) are displayed for the six related duplexes. In these isotherms the fraction bound in the primary mode, F_{B1}, is plotted versus log [*Bam*HI]. The curves correspond to DNA I (●), DNA II (○), DNA III (■), and DNA IV (□) (depicted in Figure 6) and two 22 base pair DNAs having the sequence of module A (▲) and module B (△) from our previous study (*6*). From this plot it is readily apparent that binding to the 22 base pair duplexes is stronger than to the 40 base pair duplexes. (b) Secondary mode binding isotherms for the four 40 base pair DNAs and two 22 base pair duplexes. Binding curves for the secondary mode (second gel-shifted material) are displayed for the six related duplexes. In these isotherms the fraction bound in the secondary mode, F_{B2}, is plotted versus log [*Bam*HI]. The curves correspond to DNA I (●), DNA II (○), DNA III (■), and DNA IV (□) (depicted in Figure 6)and DNAs with the module A (▲) and module B (△) sequences from our previous study (*6*). It is evident that the binding curves for the 22 base pair DNAs are far more resolved than for the 40 base pair duplexes.

Comparisons of equilibrium binding by *Bam*HI for the 22 base pair duplexes with the sequences of modules A and B and the four 40 base pair duplexes revealed significant differences in their binding isotherms. Figure 7a shows the binding isotherms for the six duplexes. Two of these (▲, and △) are for the 22 base pair duplexes with the module A and B sequences. The remaining curves are those also shown in Figure 6a. It is readily apparent from this plot that binding of *Bam*HI to the 22 base pair duplexes is much stronger than to the 40 base pair duplexes. The module A and module B DNAs have dissociation constants of 0.5×10^{-9} M and 12.0×10^{-9} M, respectively. This is 35 (module A duplex compared to DNA II) to 128 (module B duplex compared to DNA IV) times greater than for the 40 base pair duplexes. Another difference is the concentration range over which total binding occurs. For the 22 base pair DNAs the *Bam*HI concentration range for binding was approximately 5×10^{-11} M to 1×10^{-7} M, or about 3.5 log units, whereas for DNAs I−IV binding occurs over the range of 6.0×10^{-8} to 1.0×10^{-6} M, or about 1.5 log units.

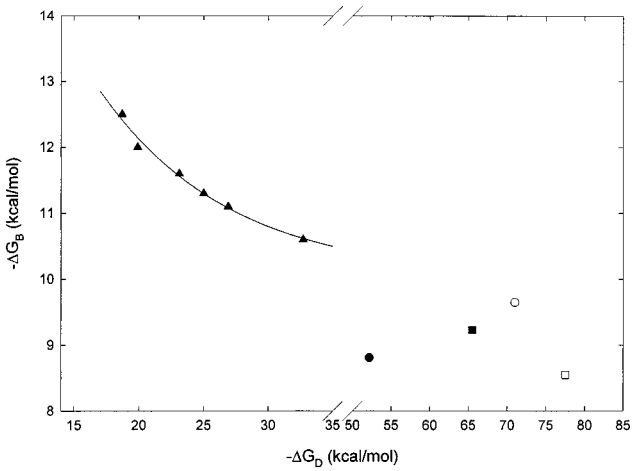

FIGURE 8: Free energy of melting at 20 °C determined by DSC, $-\Delta G_D$, is plotted versus the free energy at 20 °C of binding in the primary mode, $-\Delta G_B$, for the six 22 base pair DNAs of our previous study (▲; *6*) and the four 40 base pair DNA duplexes of the present study (○, ●, ■, and □). From this plot a clear correlation between binding and stability exists for the six 22 base pair DNA duplexes, but does not exist for the 40 base pair DNA duplexes.

The appearance of second shifted material (the secondary binding mode) was observed for both the 22 and 40 base pair duplexes. However, there are significant differences in the behaviors of the two sets of molecules. For the 22 base pair duplexes the primary and secondary binding modes of binding did not overlap; both modes were fully resolved and could be fit to logistic functions. More importantly, the correlation between K_{B2} and ΔG_D seen for the secondary mode of binding of the 22 base pair duplexes is not observed for the 40 base pair duplexes. Figure 7b shows the secondary binding data for the 22 base pair module A and B duplexes (▲ and △) along with those for the four 40 base pair DNAs. It is evident for the 22 base pair DNAs that the curves are well-resolved, yielding distinct K_{B2} values. The secondary binding mode for the 40 base pair duplexes occurs at higher *Bam*HI concentrations and over a much smaller concentration range (less than 1 log unit). No correlation between K_{B2} and ΔG_D for the 40 base pair duplexes could be discerned.

A major difference between the 22 and 40 base pair duplexes is the lack of correlation between stability and primary binding. In the study of the six 22 base pair duplexes, a plot of $-\Delta G_B$ versus $-\Delta G_D$ revealed a decaying exponential relationship between equilibrium binding and DNA stability (as measured by DSC experiments). This correlation is recounted in Figure 8, where $-\Delta G_B$ is plotted versus $-\Delta G_D$ for the six 22 base pair DNA duplexes and the four 40 base pair DNAs. For the longer DNAs a clear correlation between DNA stability and enzyme binding does not emerge. This might have been expected since it was shown previously that as the length of a set of duplexes was extended, variations in binding/reactivity directly attributable to flanking sequence stability decreased (*1,6*). However, all of the duplexes we have studied before now were much shorter and ranged from 12 to 22 base pairs. One obvious explanation for these findings relies on the assumption that structural deformations of the duplex DNA around the binding site are

required for binding. If so, then such deformations occur with greater facility in short DNA duplexes, thermodynamically involve a substantial portion of the total stability of the DNA, and thus are a strong function of the melting stability of the flanking sequence. However, as more sequence is added to the ends the required deformations become (are) harder to obtain, binding is weaker, and the effect of flanking sequence stability on the binding is less apparent because the thermodynamics of the deformation are a much smaller proportion of the total stability of the DNA.

Thus, the absence of a correlation between stability and binding (for the 40 base pair duplexes) suggests that such relationships may only be measurable in relatively short duplex oligomeric systems. If such effects do persist in longer DNAs, they do not manifest in our experiments to a significant extent to be effectively probed with this type of gel-shift assay. Results for the 40 base pair DNA duplexes suggest that equilibrium binding in the absence of Mg^{2+}, as detected in our assay, is not affected in a measurable way by the thermodynamic stability of the flanking sequence. Apparently in a longer sequence environment binding affinity of *Bam*HI cannot be related simply to sequence-dependent stability of the flanking sequences.

ACKNOWLEDGMENT

We thank Dr. Ira Schildkraut of New England Biolabs for his generous gift of the highly pure, magnesium-free *Bam*HI enzyme used in these studies. We also thank Dr. Peter V. Riccelli for helpful comments on the manuscript. Portions of this work appeared in the Ph.D. Thesis of Peter M. Vallone, Department of Chemistry, University of Illinois, Chicago, August 1999.

REFERENCES

1. Benight, A. S., Gallo, F. J., Paner, T. M., Bishop, K. D., Faldesz, B. D., and Lane, M. J. (1995) *Adv. Biophys. Chem. 55*, 1−55.
2. Alves, J., Pingoud, A., Haupt, W., Langowski, J., Peters, F., Maass, G., and Wolff, C. (1984) *Eur. J. Biochem. 140*, 83−92.
3. Armstrong, K., and Bauer, W. K. (1982) *Nucleic Acids Res. 10*, 993−1007.
4. Drew, H. R., and Travers, A. A. (1985) *Nucleic Acids Res. 13*, 4445−4467.
5. Goldstein, K., Thomas, M., and Davis, R. (1975) *Virology 66*, 420−427.
6. Riccelli, P. V., Vallone, P. M., Kashin, I., Faldasz, B. D., Lane, M. J., and Benight, A. S. (1999) *Biochemistry 38*, 11197−11208.
7. Taylor, J. D., and Halford, S. E. (1992) *Biochemistry 31*, 90−97.
8. Cantor, C. R., Warshaw, M. M., and Shapiro, H. (1970) *Biopolymers 9*, 1059−1077.
9. Vallone, P. M., Paner, T. M., Hilario, J., Lane, M. J., Faldasz, B. D., and Benight, A. S. (1999) *Biopolymers 50*, 425−442.
10. Marky, L. A., and Breslauer, K. J. (1987) *Biopolymers 26*, 1601−20.
11. Roger, A., and Norden, B. (1997) *Circular Dichroism and Linear Dichroism*, Oxford University Press, Oxford (England) and New York.
12. Johnson, W. C., Jr. (1985) *Methods Biochem. Anal. 31*, 62−163.
13. Amaratunga, M., Pancoska, P., Paner, T. M., and Benight, A. S. (1990) *Nucleic Acids Res 18*, 577−582.
14. Carey, J. (1991) *Methods Enzymol. 208*, 103−117.
15. Fried, M. G. (1989) *Electrophoresis 10*, 366−376.
16. Fried, M. G., and Crothers, D. M. (1981) *Nucleic Acids Res. 9*, 6505−6525.
17. Garner, M. M., and Revzin, A. (1981) *Nucleic Acids Res. 9*, 3047−3060.
18. Jack, W. E., Greenough, L., Dorner, L. F., Xu, S. Y., Strzelecka, T., Aggarwal, A. K., and Schildkraut, I. (1991) *Nucleic Acids Res. 19*, 1825−1829.
19. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1992) *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
20. Senear, D. F., and Brenowitz, M. (1991) *J. Biol. Chem. 266*, 13661−13671.
21. Holbrook, J. A., Capp, M. W. Saecker, R. M. and Record, M. T., Jr. (1999) *Biochemistry 38*, 8409−8422.
22. Chalikian, T. V., Völker, J., Plum, G. E., Breslauer, K. J. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 7853−7858.
23. Rouzina, I., and Bloomfield, V. A. (1999) *Biophys. J. 77*, 3242−3251.
24. Rouzina, I., and Bloomfield, V. A. (1999)*Biophys. J. 77*, 3252−3255.
25. Jelesarov, I., Crane-Robinson, C., and Privalov, P. L. (1999) *J. Mol. Biol. 294*, 981−995.
26. Privalov, P. L., Jelesarov, I., Read, C. M., Dragan, A. I. and Crane-Robinson, C. (1999) *J. Mol. Biol. 294*, 997−1013.
27. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1988) *Numerical Recipes*, Cambridge University Press, Cambridge (England) and New York.
28. Sheardy, R. D., Suh, D., Kurzinsky, R., Doktycz, M. J., Benight, A. S., and Chaires, J. B. (1993) *J. Mol. Biol. 231*, 475−488.
29. Aida, M. (1988) *J. Theor. Biol. 130*, 327−335.
30. Allawi, H. T., and SantaLucia, J., Jr. (1997) *Biochemistry 36*, 10581−10594.
31. Breslauer, K. J., Frank, R., Blocker, H., and Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746−3750.
32. Delcourt, S. G., and Blake, R. D. (1991) *J. Biol. Chem. 266*, 15160−15169.
33. Doktycz, M. J., Goldstein, R. F., Paner, T. M., Gallo, F. J., and Benight, A. S. (1992) *Biopolymers 32*, 849−864.
34. Gotoh, O., and Tagashira, Y. (1981) *Biopolymers 20*, 1033−1042.
35. McCampbell, C. R., Wartell, R. M., and Plaskon, R. R. (1989) *Biopolymers 28*, 1745−1758.
36. SantaLucia, J., Allawi, H. T., and Seneviratne, P. A. (1996) *Biochemistry 35*, 3555−3562.
37. Sugimoto, N., Nakano, S., Yoneyama, M., and Honda, K. (1996) *Nucleic Acids Res. 24*, 4501−4505.
38. Vologodskii, A. V., Amirikyan, B. R., Lyubchenko, Y. L., and Frank-Kamenetskii, M. D. (1984) *J. Biomol. Struct. Dyn. 2*, 131−148.
39. Wartell, R. M., and Benight, A. S. (1985) *Phys. Rep. 126*, 67−107.
40. Owczarzy, R., Vallone, P. M., Gallo, F. J., Paner, T. M., Lane, M. J., and Benight, A. S. (1997) *Biopolymers 44*, 217−239.
41. SantaLucia, J., Jr. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 1460−1465.
42. George, J., Nardone, G., and Chirikjian, J. G. (1985) *J. Biol. Chem. 260*, 14387−14392.
43. Grabowski, G., Jeltsch, A., Wolfes, H., Maass, G., and Alves, J. (1995) *Gene 157*, 113−118.
44. Xu, S. Y., and Schildkraut, I. (1991) *J. Bacteriol. 173*, 5030−5035.
45. Xu, S. Y., and Schildkraut, I. (1991) *J. Biol. Chem. 266*, 4425−4429.
46. Strzelecka, T., Newman, M., Dorner, L. F., Knott, R., Schildkraut, I., and Aggarwal, A. K. (1994) *J. Mol. Biol. 239*, 430−432.

BI000326K