Thermodynamic, Spectroscopic, and Equilibrium Binding Studies of DNA Sequence Context Effects in Six 22-Base Pair Deoxyoligonucleotides

Peter V. Riccelli,[‡,§] Peter M. Vallone,[‡] Irina Kashin,[§] Brian D. Faldasz,[§] Michael J. Lane,[§,||] and Albert S. Benight*[,‡]

Department of Chemistry, Room 4500, 845 West Taylor Street, University of Illinois, Chicago, Illinois 60607, Tm Technologies, Inc., 82 Cummings Park, Woburn, Massachusetts 01801, and Departments of Medicine and Microbiology & Immunology, State University of New York-Health Science Center at Syracuse, Syracuse, New York 13207

Received February 24, 1999; Revised Manuscript Received May 20, 1999

ABSTRACT: Effects of different end sequences on stability, circular dichroism spectra (CD), and enzyme binding properties were investigated for six 22-base pair, non-self-complementary duplex DNA oligomers. The center sequences of these deoxyoligonucleotides have 8−14 base pairs in common and are flanked on both sides by sequences differing in context and A-T 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 van't Hoff analysis of optical melting curves collected as a function of DNA concentration, assuming a two-state melting transition. Melting free energies (20 °C) of the six DNAs evaluated from DSC experiments ranged from −18.7 to −32.7 kcal/mol. van't Hoff estimates of the free energies ranged from −18.5 to −48.0 kcal/mol. With either method, the trends in free energy as a function of sequence were identical. Equilibrium binding by *Bam*HI restriction endonuclease to the 22-base pair DNAs was also investigated. The central eight base pairs of all six molecules, 5′-A-GGATCC-A-3′, contained a *Bam*HI recognition sequence bounded by A-T base pairs. Magnesium free binding assays were performed by titering *Bam*HI against a constant concentration of each of the deoxyoligonucleotide substrates and analyzing reaction products by gel retardation. Binding isotherms of the total amount of bound DNA versus protein concentration were constructed which provided semiquantitative estimates of the equilibrium dissociation constants for dissociation of *Bam*HI from the six DNA oligomers. Dissociation constants ranged from 0.5×10^{-9} to 12.0×10^{-9} M with corresponding binding free energies of −12.5 to −10.6 (±0.1) kcal/mol. An inverse relationship is found when binding and stability are compared.

Sequence-dependent annealing of two complementary DNA strands is the central feature of DNA hybridization reactions. Such sequence-dependent reactions have applications in many areas of DNA research. The selection and design of probe and primer sequences for optimum performance of these reactions requires an accurate prediction of the DNA sequence-dependent melting stability of the duplex to be formed by hybridization. For this reason, a substantial effort has been undertaken with the goal of obtaining the ability to make accurate calculations of DNA duplex stability from base pair sequences (*1*−*3*). Oligonucleotide probes and primers are generally between 20 and 25 base pairs long. Melting of short duplexes of this length is assumed to be a relatively simple process, i.e., a two-state process. Exceptions in which short DNA duplexes display significant departures from two-state behavior abound (more than commonly acknowledged or appreciated) (*1*, *2*).

Although the understanding of these sequence-related departures from model-dependent behaviors is not complete, sequence-dependent end effects on overall duplex stability may be an important factor. In this paper, we explore the effects of sequence context on melting, CD spectra, and enzyme binding for six 22-base pair duplex DNAs. Results show that sequence-dependent end effects not only are determined by end sequences themselves but also depend on the sequence environment adjoining the end sequences. These sequence context effects are manifested in altered sequence-dependent nucleation of the duplexes. Effects of sequence context on their CD spectra are also seen. In addition, results indicate that equilibrium binding of the *Bam*HI restriction enzyme depends on the sequence content and the context of duplex base pairs.

MATERIALS AND METHODS

DNA Samples. DNA oligomer strands that form the duplexes in Figure 1 were synthetically prepared. Each DNA has a central eight-base pair sequence 5′-A-*G*-*G*-*A*-*T*-*C*-*C*-A-3′. Italicized base pairs in Figure 1 denote the six base pairs that comprise the cognate recognition sequence for the *Bam*HI restriction endonuclease. In addition to the central eight base pairs common to all molecules, two subsets can be denoted according to their common central 14-base pair

* To whom correspondence should be addressed.
‡ University of Illinois.
§ Tm Technologies, Inc.
|| State University of New York-Health Science Center at Syracuse.

10.1021/bi9904407 CCC: $18.00
Published on Web 08/05/1999

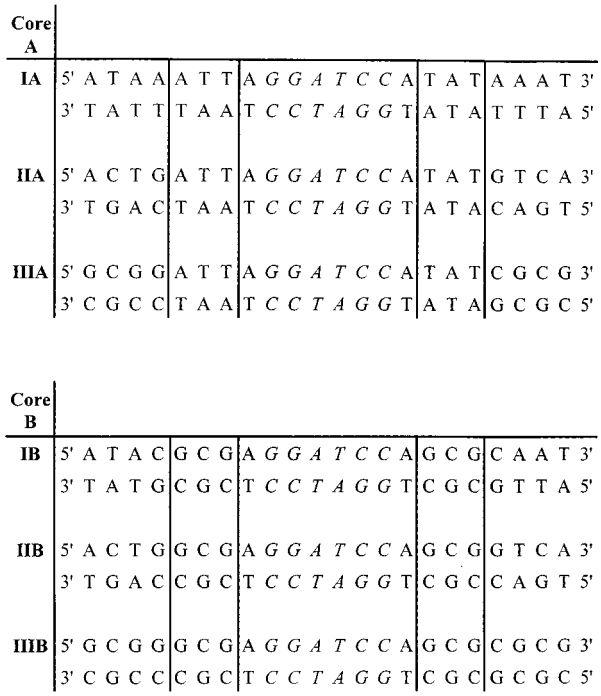

FIGURE 1: Sequences of the 22-base pair duplex DNA molecules. Sequences are designated I−III according to their common 14-base pair modules A and B. The eight-base pair core contains the cognate site (italic) for the *Bam*HI restriction enzyme, flanked on both sides by A-T residues. The remaining three base pairs adjoining the central common sequences define modules A and B as indicated. The same end sequences flank the module A and module B sequences.

module sequences. These are denoted as module A and module B sequences in Figure 1. The molecules of module A have the same central 14 base pairs in which the universally common eight-base pair sequence is flanked on the left by the sequence 5′-A-T-T-3′ and on the right by 5′-T-A-T-3′. The common central sequence is surrounded on both ends by different arrangements of A-T (T-A) or G-C (C-G) type base pairs. The module B molecules have 14 central base pairs that are different from those of module A with the universally common eight-base pair sequence flanked on the left and the right by the sequence 5′-G-C-G-3′. As for the module A molecules, the 14-base pair common sequence of module B is surrounded on both sides by different arrangements of A-T (T-A) or G-C (C-G) type base pairs. Inspection of the sequences in Figure 1 shows that sequences flanking the eight common central base pairs vary in composition from purely A-T (IA) or G-C (IIIB) type base pairs to three contiguous A-T (IIA and IIIA) or G-C (IB and IIIB) base pairs followed by end sequences of a mixture of G-C type base pairs. These oligomer sequences were designed with the intent of varying the sequence-dependent stability of the sequences flanking the central sequence in an incremental manner so as to span the total range of sequence-dependent melting stabilities that are available. The *Bam*HI binding site was engineered in the centers of these DNAs to allow a systematic evaluation of the effects of flanking sequence stability on *Bam*HI binding. In addition, a 17-base pair duplex DNA (DNA IV, not shown in Figure 1) with the sequence 5′-GAGATACATATATATC-GAT-3′ was prepared and served as a control in electrophoretic gel shift studies. Note that the sequence of DNA IV does not contain the cognate binding sequence for *Bam*HI.

Individual DNA single strands used to make the duplexes were synthesized on an Applied Biosystems 380B automated DNA synthesizer using standard phosphoramidite chemistry (*4*). Strands were deprotected in ammonia and exhaustively dialyzed against Nanopure water for at least 3 days, and then purified to homogeneity by electrophoresis on 20% preparative polyacrylamide gels containing 8 M urea run at 50 °C. Gels were back shadowed using a short-wave UV lamp (254 nm), while appropriate bands were sliced out of the gel. DNAs were electroeluted from gel slices and passed through a 0.22 μm filter. To remove dissolved urea, acrylamide, and other buffering salts, DNAs were dialyzed against several high- and low-salt solutions [100 or 1 mM NaCl and 1 mM Na_2EDTA (pH 7.3)] and finally versus Nanopure water. These purified and desalted strands were vacuum-dried on a speed vac concentrator and stored at −20 °C. Single-strand concentrations were estimated using extinction coefficients at 260 nm determined by the nearest-neighbor method (*5*).

For each duplex that was prepared, 2 pmol of one strand was radiolabeled with [γ-^{32}P]ATP (6000 Ci/mmol, New England Nuclear) using T4 polynucleotide kinase (New England Biolabs) as described previously (*6*). Reactions were optimized to label approximately 90% of the DNA by using a molar concentration of [γ-^{32}P]ATP that was 10 times higher than the DNA concentration in a 15 μL volume. Labeled strands were purified electrophoretically as described above. Bands corresponding to labeled DNA strands were excised from the gel and recovered by the crush and soak method. Gel particles were removed by passing the DNA oligomers over columns of silanized glass wool. The DNAs were desalted and transferred to the assay (and melting) buffer [105 mM sodium chloride, 10 mM sodium cacodylate, and 1 mM Na_2EDTA (pH 7.0)] by elution through Bio-gel P6 spin columns (Bio-Rad) that had been equilibrated in assay buffer. Specific activities of recovered labeled DNA strands were determined by counting in 5 mL of Sigma-Fluor scintillation cocktail fluid on a Beckman scintillation counter. Activities were typically around 3.5×10^5 cpm/μL. Total labeled strand concentrations were estimated by titrating equal amounts of radiolabeled strands with increasing amounts of unlabeled complement strands at known concentrations and analysis of the annealed products on 15% polyacrylamide gels. Strands were used in experiments within 1 week of being labeled and purified.

BamHI Protein. *Bam*HI protein was isolated as described previously (*7*). Analysis of the protein by SDS−polyacrylamide gel electrophoresis revealed it to be greater than 99.9% pure. The concentrated protein solution (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). Enzyme activity was reported to be 10^6 units/mg (personal communication from I. Schildkraut, New England Biolabs). Stock solutions of the concentrated protein were divided into 25 μL aliquots and stored at −70 °C until they were used in assays. Stock protein solutions for binding assays were obtained by serial dilution in 10 mM Tris-HCl (pH 7.0), 1 mM dithiothreitol, 0.1 mM Na_2EDTA, and 10% glycerol, to concentrations ranging from 15×10^{-5} to 15×10^{-17} M, and prepared fresh prior to each binding assay. Molar quantities used in titrations were obtained from the known protein concentration in milligrams per milliliter.

Optical Melting Studies. For optical melting studies, DNA strands were rehydrated in melting buffer [105 mM sodium chloride, 10 mM sodium cacodylate, and 1 mM Na_2EDTA (pH 7.0)]. Duplex stock solutions of the sequences in Figure 1 were prepared by mixing complementary strands in 1:1 molar ratios and heating and cooling as described above. Complete duplex formation was verified by analysis on 20% polyacrylamide gels run at 18 °C.

Optical melting experiments were performed on an HP8452 single-beam diode array spectrophotometer. DNAs were melted over the strand concentration range of 0.5−70 $\times 10^{-6}$ M. Melts were performed in either 1.0 or 0.1 cm quartz cuvettes. Prior to melting, samples were filtered through 0.45 μm filters and degassed by bubbling with a fine stream of helium for at least 30 min. Three drops of mineral oil were layered on the surface, and a Teflon-coated temperature probe was inserted directly into the sample. The cuvette was sealed with Teflon tape. Samples were placed in a thermostated cell holder and equilibrated at the starting temperature for approximately 30 min. Melting experiments performed at higher concentrations, where the optical density exceeded 0.8 OD/mL, were performed using 0.1 cm path length cuvettes. For these experiments, 800 μL of DNA solution was filtered into a buffer-rinsed Eppendorf tube and carefully pipetted directly into the cuvette. The sample was degassed by incubation, unsealed at 20 °C. Cuvettes were sealed with Teflon tape, inserted into a thermostated cell holder, and equilibrated at the starting temperature. At least four forward and reverse curves were collected for each concentration at 268 nm at a heating rate of 35 °C/h.

Absorbance versus temperature curves were normalized to upper and lower linear baselines to obtain fraction of broken base pairs (θ) versus temperature curves. The transition temperature, t_m, was evaluated at $\theta = 0.5$. The reproducibility of experimentally determined t_m values was ±0.5 °C. Melting data were analyzed assuming a two-state model for the duplex to single-strands transition. From linear fits of van't Hoff plots, constructed by plotting $1/t_m$ versus $\ln(C_T/4)$, the van't Hoff enthalpy, ΔH_{vH}, and entropy, ΔS_{vH}, were evaluated. From these quantities, the free energy, ΔG_{vH}, was determined.

Differential Scanning Calorimetry. DNA duplexes were prepared for calorimetric melting experiments by exhaustive dialysis against melting buffer [105 mM NaCl, 10 mM sodium cacodylate, and 0.1 mM Na_2EDTA (pH 7.0)]. Excess heat capacity changes, ΔC_p^{ex}, of DNA solutions were measured as a function of temperature on a Microcal MC-2 ultrasensitive differential scanning calorimeter (Northampton, MA). Prior to collection of melting data, the calorimeter was calibrated with a known standard electrical pulse. For melting experiments, sample concentrations ranged from 50 to 134 μM (at least 30 OD/mL) in a total volume of approximately 2 mL. Buffer versus buffer cells (reference and sample cell both containing melting buffer) were scanned to provide baselines of the instrumental variation between the cells of the calorimeter. Then DNA versus buffer scans were collected. For each DNA sample, 8−10 scans were collected from 4 to 100 °C at a rate of 45 °C/h. After collection of DNA versus buffer scans for a particular sample, the calorimeter was cleaned and another buffer versus buffer scan was taken before the next DNA sample was added and its scan versus buffer was collected. Resulting DNA versus buffer scans collected for each sample were averaged and buffer versus buffer baselines subtracted to obtain average ΔC_p^{ex} versus T curves for each sample. These baseline-corrected curves were normalized to total DNA concentration, and a smooth cubic function baseline was fit to pretransition and post-transition regions of each curve using the Origin software package. The cubic fitted baseline was subtracted from the actual ΔC_p^{ex} versus T curve; thus, $\Delta C_p^{ex}(100\ °C) - \Delta C_p^{ex}(4\ °C) = 0$. The integrated area under each baseline-corrected ΔC_p^{ex} versus T curve provided the calorimetric transition enthalpy

$$\Delta H_{cal} = \int \Delta C_p^{ex}\, dT \quad (1)$$

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

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

Circular Dichroism (CD) Measurements. Circular dichroism (CD) spectra were recorded on a Jasco J-600 spectropolarimeter. For CD measurements, DNA sample solutions were prepared by diluting a deoxyoligonucleotide solution to a final concentration of 20 μM (duplex) in melting buffer and filtering into a 0.1 cm path length quartz cell. Samples were equilibrated for at least 30 min at 22 °C. DNA spectra were measured at a constant temperature of 22 °C maintained by a thermostated sample holder. For each DNA sample, at least five scans of the spectrum were collected over the wavelength range of 200−320 nm at a scanning rate of 10 nm/min. The averages of multiple scans were used for analysis. The scan of the buffer alone recorded at 22 °C was subtracted from the average scan for each DNA duplex. Data were collected in units of millidegrees versus wavelength and normalized to total DNA concentration. DNA concentrations were determined from the optical density at 260 nm using extinction coefficients estimated from the nearest-neighbor method (*5*). Resulting CD spectra are reported as plots of $\Delta\epsilon$ ($M^{-1}\ cm^{-1}$) versus wavelength.

Measured CD spectra collected for the six DNA molecules in Figure 1 were also analyzed by factor analysis using singular-value decomposition (SVD) (*8*−*10*). Application of this analysis provides a means for determining in a least-squares sense the minimum number of basis spectra with significant weight that can be linearly combined to fit the collection of measured CD spectra within certain limits. Basis spectra and their linear coefficients were evaluated for each DNA from SVD analysis of the set of collected CD spectra. The analytical procedure was essentially that previously described (*8*).

Binding Studies. Studies of the equilibrium binding of *Bam*HI to the DNA substrates displayed in Figure 1 were performed in the absence of Mg^{2+} where enzyme cleavage activity is eliminated, but binding activity is retained (*11*). Our intent was to use the *Bam*HI system to probe the sequence context effects on equilibrium binding to the deoxyoligonucleotides. The *Bam*HI restriction enzyme was chosen because of its high sequence specificity. Another reason for choosing *Bam*HI was the fact that the high-quality ultrapure enzyme was available in sufficient quantities for binding studies. By monitoring binding in the absence of

cleavage activity, we could examine sequence effects on the formation of enzyme−DNA complexes, not cleaved DNA products. For our purposes of probing sequence context effects, investigations of the kinetics of enzyme cleavage were beyond the scope of this study.

The mixtures for reactions performed in 15 μL volumes contained 1×10^{-10} M DNA duplexes ($2-5 \times 10^3$ cpm) and *Bam*HI dimer concentrations ranging from 1×10^{-17} to 1×10^{-5} M. Deoxyoligonucleotide substrates were prepared by mixing unlabeled DNA strands with a slight molar excess of their labeled complementary strands in assay buffer [105 mM sodium chloride, 10 mM sodium cacodylate, and 1 mM Na_2EDTA (pH 7.0)] and heating to 90 °C for 4 min and annealed by cooling at room temperature for at least 2 h. *Bam*HI was added to pre-equilibrated mixtures of DNA in assay buffer and incubated for 45 min at 20 °C in a PTC-100 thermocycler (M. J. Research). Binding reactions were analyzed in gel retardation assays. Samples were loaded on 15% polyacrylamide gels that had been pre-electrophoresed at 200 V at 18 °C for at least 4 h. Electrophoresis was carried out at 210 V for 4.5 h in 1× TBE (pH 8.2, 18 °C). The gel temperature was maintained by continually circulating water through heat exchange tubes in the gel buffer tank. Autoradiograms were obtained by exposing gels to Fuji RX X-ray film with intensifying screens for 1−2 days at −70 °C.

Autoradiograms were developed and scanned on a MicroTek IIXE scan maker. Image analysis was performed using NIH 1.52 software. Intensity peaks from scanned gels were smoothed, and background intensities were subtracted. Pixel densities of all shifted material, taken to correspond to bound DNA, were divided by total pixel density of all the bands in that lane, shifted and unshifted, to obtain the individual fractions of shifted, labeled material. For each DNA substrate, at least three separate experiments were conducted over the entire titration range at various increments of enzyme concentration. Variations from experiment to experiment were on average ≤12%. Confirmation of the accuracy of measured fractions estimated from integration of the pixel densities was obtained by randomly selecting gels, cutting out the bands corresponding to all the labeled material, and determining the actual radioactivity in each of these bands by counting in a scintillation counter. Values of bound fractions obtained in this manner yielded dissociation constants that varied less than 10% from those determined from pixel densities.

Analysis of Binding Isotherms. A formal expression for analyzing plots of the fraction of bound duplex, F_b, versus log[*Bam*HI] was derived from the simplest model that could quantitatively be fit to the experimental data. This model assumes binding to discrete noninteracting sites and does not consider binding cooperativity. Consider the dissociation and binding reaction between the duplex DNA oligomer (D) and restriction enzyme (E),

$$\mathrm{DE} \xrightarrow{K_D} \mathrm{D} + \mathrm{E} \tag{3}$$

The dissociation equilibrium constant K_D is defined as [D][E]/[DE]. In terms of total concentrations of enzyme dimers, E_o, and duplex DNA, D_o, and n enzyme dimers binding per duplex, the following relationships exist.

$$[\mathrm{D}] = D_o - [\mathrm{DE}] \tag{4a}$$

$$[\mathrm{E}] = E_o - n[\mathrm{DE}] \tag{4b}$$

Using these expressions and the definition of K_D yields the following quadratic form for F_b in terms of D_o, E_o, n, and K_D

$$F_b = \{E_o + nD_o + K_D - [(E_o + nD_o + K_D)^2 - 4nD_oE_o]^{1/2}\}/2nD_o \tag{5}$$

This expression was employed to determine values of K_D empirically, as adjustable parameters in fits of experimentally derived binding isotherms of F_b versus log[*Bam*HI]. Two bound species were observed. The first mode that appears as a band with retarded mobility at low-concentration regions of the enzyme titration was presumed to correspond to binding at the central specific site. For analysis of these binding curves, an n of 1 was assumed in eq 5. At higher protein concentrations in the titration, a second bound species with even greater retarded mobility was observed. For analysis of binding curves for this secondary binding mode, an n of 2 was assumed in eq 5.

RESULTS AND DISCUSSION

DNA Stability from Optical Melting Curves. Ultraviolet absorbance versus temperature melting curves were collected for the DNAs of Figure 1 as a function of total strand concentration, C_T, from 0.5 to 70 μM. The averages of three such curves measured for each DNA at a strand concentration of 5 μM are shown in Figure 2, where relative absorbance increases [absorbance(T)/initial absorbance at 268 nm] are plotted versus temperature. The individual strands used to form the duplexes in Figure 1 did not show any evidence for intramolecular structures. All strands had identical gel mobilities on polyacrylamide gels run under native conditions (15 °C). In addition, mixing of the strands and their complement at a 1:1 molar ratio resulted in single products. Excess single strands in the presence of duplexed strands ran to the same relative position on gels as the single strands alone. Thus, the predominant form of the sequences in Figure 1 is the duplex. The comparisons in Figure 2 clearly show, at the same strand concentration, differences in t_m values of the DNA sequences generally increase with increasing differences in their %GC. At the DNA concentration where the curves in Figure 2 were acquired (5 μM), t_m values of the duplexes varied over a 35 °C range. Figure 2A is the data for the module A molecules IA (○), IIA (△), and IIIA (□). The sequences of these DNAs have 18.2% (IA), 36.4% (IIA), and 54.4% (IIIA) G-C type base pairs. Figure 2B shows melting data for the module B molecules IB (●), IIB (▲), and IIIB (■). The %GC values of these sequences are 54.4% (IB), 63.6% (IIB), and 81.8% (IIIB). Comparison of the melting curves for the module A molecules based on their differences in %GC shows the expected increase in stability with increases in the %GC of sequences flanking the common central sequence. The module B molecules IB (●) and IIB (▲) have the same stability, although IIB has one more G-C base pair than IB. In addition, differences exist in the melting curves for molecules IIIA and IB (□ and ●, respectively). These sequences have the same %GC

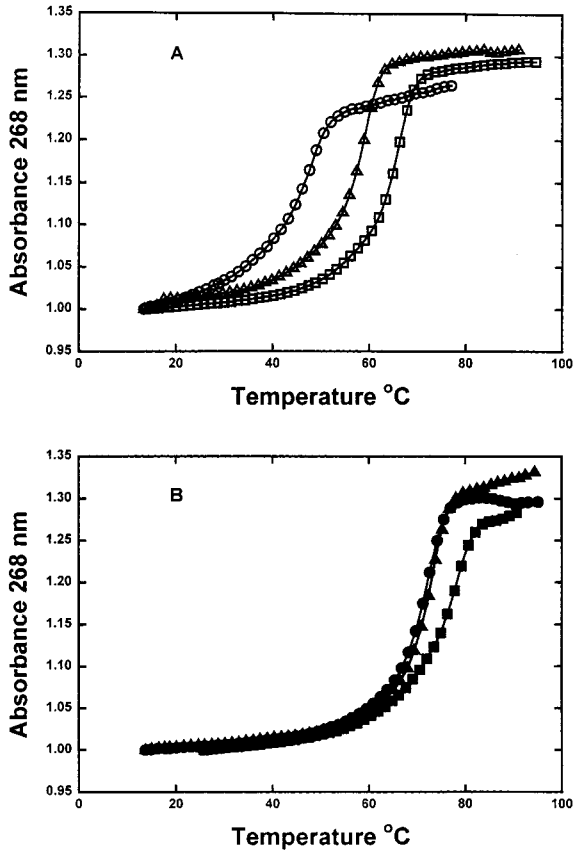

FIGURE 2: Duplex optical melting curves. Relative absorbance at 268 nm vs temperature melting curves collected as a function of DNA concentration for the six deoxyoligonucleotide sequences of Figure 1. Representative curves shown here were for a single DNA strand concentration of 5×10^{-6} M. (A) Module molecules IA (○), IIA (△), and IIIA (□). (B) Module B molecules IB (●), IIB (▲), and IIIB (■).

(54.5%) and differ only in the relative order of their A-T and G-C type base pairs, but their melting curves do not overlap exactly. Thus, not only the sequence content (%GC) but also the order, or context, of the base pairs determines the melting stability of the DNAs.

van't Hoff plots of $1/t_m$ versus $\ln(C_T/4)$ were constructed for the six duplexes from the concentration dependence of the t_m values. These plots are shown in Figure 3. Plots for the module A molecules (white symbols) and module B molecules (black symbols) 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}. The underlying assumption when applying this graphical evaluation of the thermodynamic parameters is that the DNA melting transition occurs in a two-state manner. These graphically evaluated thermodynamic parameters are summarized in Table 1. The values of ΔH_{vH} range from −123.6 kcal/mol for DNA IA to −248.6 kcal/mol for DNA IIIB, but the remaining values for ΔH_{vH} do not increase with increasing %GC. That is, DNAs IIA and IIB that differ considerably in %GC (36.4 vs 63.6%) have the same ΔH_{vH} values (approximately −150 kcal/mol), while DNAs IIIA and IB, which have the same %GC, have very different ΔH_{vH} values (−124.2 vs −157.6 kcal/mol). Similarly, the van't Hoff entropy values, ΔS_{vH}, range from −339.8 eu for DNA IIIA to −684.2 eu for DNA IIIB.

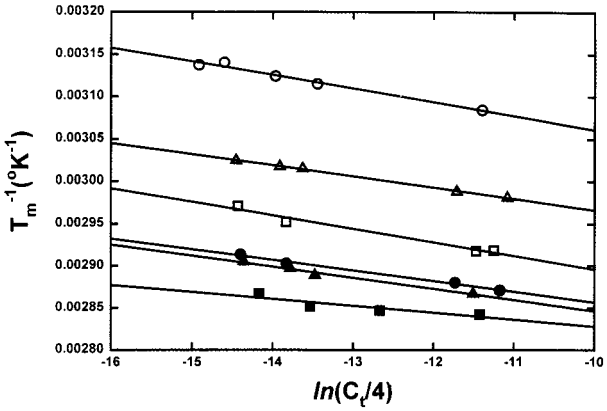

FIGURE 3: van't Hoff plots for melting of the duplexes. Results obtained for melting the duplex DNAs of Figure 1 as a function of DNA concentration. Values of t_m as a function of total strand concentration, C_t, were evaluated from melting curves such as those shown in Figure 2 and used to construct $1/t_m$ vs $\ln(C_t/4)$ plots. Analysis of these plots assuming a two-state melting transition provided evaluations of the transition enthalpy, ΔH_{vH}, and entropy, ΔS_{vH}. Plots for module A molecules IA (○), IIA (△), and IIIA (□) and plots for module B molecules IB (●), IIB (▲), and IIIB (■). The graphically evaluated thermodynamic parameters are summarized in Table 1.

Table 1

	ΔH_D (kcal/mol)		ΔS_D (eu)		ΔG_D(20 °C) (kcal/mol)	
DNA	vH	cal	vH	cal	vH	cal
IA	−123.6	−182.3	−358.5	−558.2	−18.5	−18.7
IIA	−151.0	−155.9	−427.9	−464.0	−25.6	−19.9
IIIA	−124.2	−159.8	−339.8	−466.4	−24.6	−23.1
IB	−157.6	−156.8	−430.4	−449.5	−31.4	−25.0
IIB	−150.4	−167.3	−408.3	−478.9	−30.7	−26.9
IIIB	−248.6	−187.2	−684.2	−527.2	−48.0	−32.7

DNA Stability As Determined by Differential Scanning Calorimetry. Results from DSC melting experiments are shown in Figure 4. Excess heat capacity, ΔC_p^{ex}, is plotted versus temperature for the module A and module B DNAs. The calorimetric melting curves are clearly resolved for the module A molecules, indicating a significant difference in stability due to the ends for these DNAs. In contrast, the curves for module B molecules IB and IIB almost overlay exactly, as was observed with optical melting curves of these sequences, revealing that the different end sequences of these molecules have little effect on the calorimetric melting transitions. The t_m values of these curves were determined from the temperature of the peak height maximum. Evaluations of the transition enthalpy and entropy, ΔH_{cal} and ΔS_{cal}, for each molecule were obtained from the integrated area under each curve. As shown in Table 1, comparison of the calorimetric melting thermodynamics with those obtained from the van't Hoff analysis of optical melting data for the module A and module B DNAs reveals some disagreement. In the case of DNA IA, ΔH_{vH} is lower than ΔH_{cal} by nearly 60 kcal/mol, while ΔH_{vH} for DNA IIIB is higher than ΔH_{cal} by about the same amount. Discrepancies observed in Table 1 suggest significant deviations from two-state melting behavior for most of the 22-base pair duplex oligomers. As a result of these discrepancies, the model-independent, directly measured calorimetric values, instead of thermodynamic parameters evaluated from graphical analysis of optical melting curves assuming a two-state transition, are deemed more reliable and used in the following comparisons.

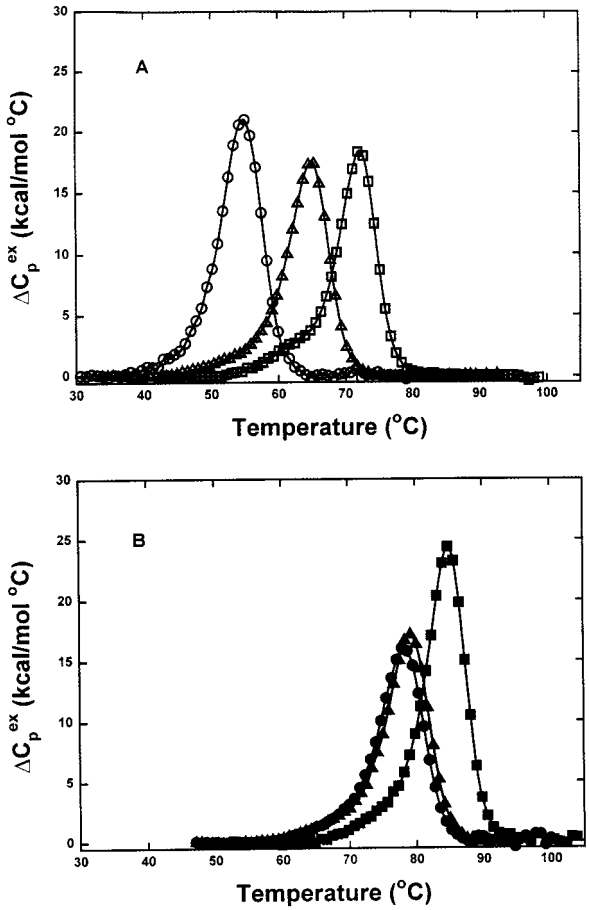

FIGURE 4: Calorimetric melting curves. Plots of the excess heat capacity, ΔC_p^{ex}, vs temperature from differential scanning calorimetry are shown for the six DNAs shown in Figure 1. (A) The scans for module A molecules IA (○), IIA (△), and IIIA (□). (B) The scans for module B molecules IB (●), IIB (▲), and IIIB (■). The thermodynamic parameters derived from these curves are summarized in Table 1.

The calorimetric t_m values for the module A and B DNAs agreed to within 1 °C with the t_m linearly extrapolated from the concentration-dependent optical melting data to the higher concentrations where DSC melting experiments were conducted. Although there are some quantitative differences between optical and calorimetric melting parameters, the qualitative differences in stability found for the DNAs by either technique are the same according to their free energy values.

Duplex Initiation Free Energies for the 22-mer DNAs Are Sequence-Dependent. The melting transition free energies for the duplex sequences of the module A and module B molecules can be calculated directly from any of the published sets of nearest-neighbor (n-n) sequence-dependent stability parameters (*2*, *3*, *12*−*20*). A number of these sets have been reported and recently reviewed and compared (*1*). For our calculations, we employed the n-n set evaluated from melting studies of DNA dumbbells (*14*). Comparisons of our n-n parameters and n-n parameters from other laboratories were recently presented, and a new "unified" n-n parameter set was presented (*21*). In that analysis, our n-n set, evaluated from melting analysis of DNA dumbbells and used here to calculate DNA stability, was found to be in good agreement with the unified n-n set. In addition, our n-n parameters seemed to be the most appropriate because they were determined in the same Na^+ environment where the module A and module B molecules were melted. The predicted free energies, determined at 20 °C from our published n-n values (*1*), ΔG_{pred}(20 °C), for the sequences of Figure 1, are given in Table 2 and compared directly with the values measured by DSC, ΔG_{cal}. The ΔG_{cal} values include contributions from both the favorable sequence-dependent (predicted) free energies and the unfavorable energetic cost of duplex initiation. The initiation free energy ΔG_{int} is defined as $\Delta G_{cal} - \Delta G_{pred}$. These values summarized in Table 2 are plotted versus $\Delta G_{cal} = \Delta G_D$ for the module A (white symbols) and module B (black symbols) molecules in Figure 5. The same symbols denote the module A and module B molecules having the same ends. Estimated uncertainties in these values are ±0.8 kcal/mol. As Table 2 and Figure 5 clearly show, the determined values of ΔG_{int} are not the same for different module sequences, and range from 11.6 kcal/mol for DNA IA to 15.9 kcal/mol for DNA IIIA. In addition, the magnitude of the end effect depends on the adjoining sequence environment. This sequence dependence of ΔG_{int} is a likely source for the deviations from two-state melting behavior found for these DNAs. Larger differences between ΔG_{int} values for the module A molecules (white symbols) suggest end effects may be greater for the module A sequence than for the module B sequence.

Table 2

	ΔG_D(20 °C) (kcal/mol)				ΔG_D(20 °C) (kcal/mol)		
DNA	pred	cal	int	DNA	pred	cal	int
IA	−30.3	−18.7	11.6	IB	−39.5	−25.0	14.5
IIA	−34.4	−19.9	14.5	IIB	−40.7	−26.9	13.8
IIIA	−39.0	−23.1	15.9	IIIB	−46.0	−32.7	13.3

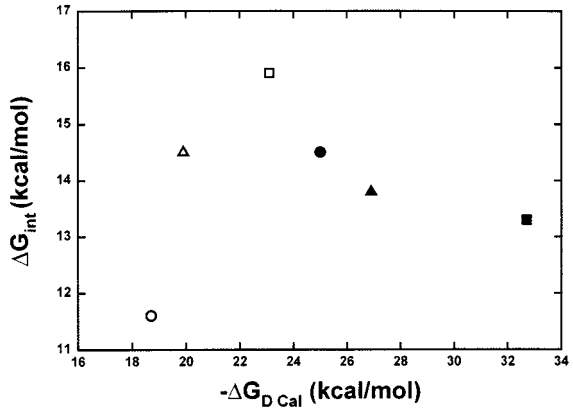

FIGURE 5: Duplex initiation free energies. Duplex initiation free energy, ΔG_{int}, determined from the difference between the calorimetrically evaluated free energy, ΔG_{Dcal}, and the free energy predicted from the sequence, ΔG_{pred}, from the nearest-neighbor stability parameter, plotted vs ΔG_{Dcal}. Data for module A molecules IA (○), IIA (△), and IIIA (□). Data for module B molecules IB (●), IIB (▲), and IIIB (■). These results show that ΔG_{int} is not constant for the deoxyoligonucleotides shown in Figure 1. Free energy values are at 20 °C.

Circular Dichroism Spectroscopy of the DNA Substrates. The CD spectra for the module A and module B DNAs collected over the wavelength range of 200−340 nm are shown in Figure 6. As in previous figures, the spectra for molecules with the two module sequences are grouped for comparison. Figure 6A shows the spectra for the module A DNAs: IA (○), IIA (△), and IIIB (□). The CD spectra for

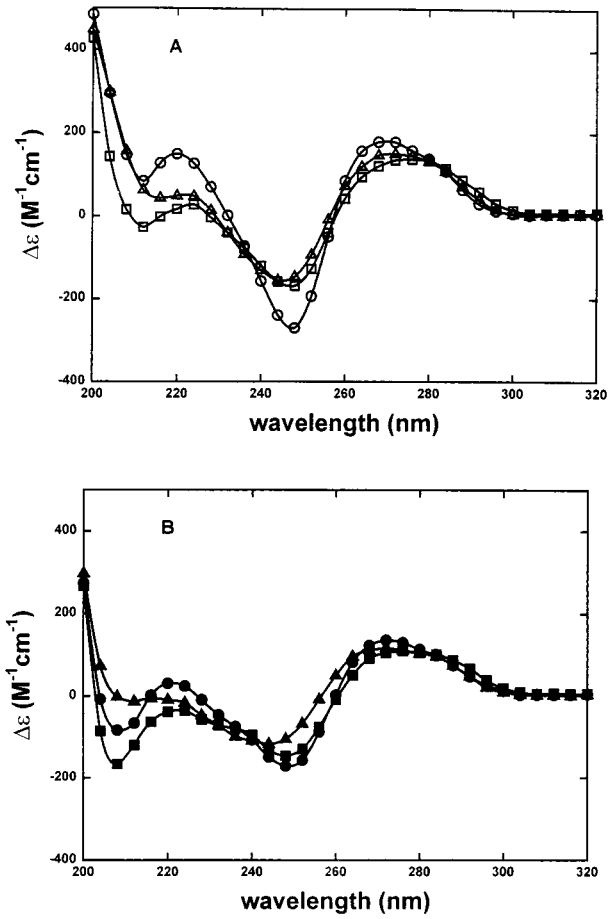

FIGURE 6: Circular dichroism spectra of the DNA substrates. CD spectra plotted as $\Delta\epsilon$ vs wavelength recorded from 200 to 320 nm for the six deoxyoligonucleotides shown in Figure 1. (A) The spectra for module A molecules IA (○), IIA (△), and IIIA (□). (B) The spectra for module B molecules IB (●), IIB (▲), and IIIB (■).

the module B DNAs, IB (●), IIB (▲), and IIIB (■), are shown in Figure 6B. Larger differences are seen in the spectra for the molecules with module A than for those with the module B sequence.

To obtain a more quantitative assessment of the distinguishable spectral components comprising the spectra in Figure 6, and their relative changes with sequence, all CD spectra were subjected to factor analysis by SVD (*8*−*10*). 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 6 revealed that the original CD spectra for all six deoxyoligonucleotides could be reconstructed from one major and two minor subspectra (results not shown). The relative weights of these subspectra were 0.84, 0.14, and 0.02. Supposedly, these subspectra (not shown) reveal the presence of spectroscopically distinguishable features of the DNAs. Changes in the coefficients of the major subspectra for module A and B DNA molecules were used as a gauge of incremental sequence-dependent variations of the CD spectra with changes in sequence.

Relationships between Melting Stability and CD Spectra. We now compare the relative magnitude of changes in the measurable parameters obtained from melting and CD analysis as a function of sequence. The calorimetric melting

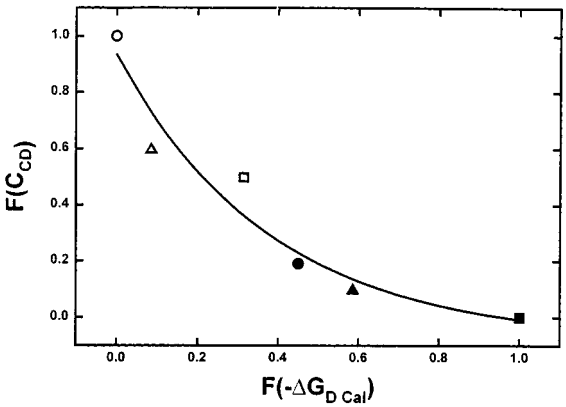

FIGURE 7: Relative comparison of the circular dichroism spectra and melting stabilities. Relative changes in the melting free energy, ΔG_{Dcal}, and the coefficient of the major CD subspectrum, C_{CD}, were normalized between the maximum differences of these quantities obtained for the DNAs shown in Figure 1. These fractional changes, $F(\Delta G_{Dcal})$, and $F(C_{CD})$ are plotted against each other. The plot shows that the relationship between $F(C_{CD})$ and $F(\Delta G_B)$ is logarithmic. The curve shown through the data corresponds to the equation $F(C_{CD}) = c \exp[-F(\Delta G_{Dcal})]$.

free energy, ΔG_{Dcal}, and the coefficient of the major CD subspectrum, C_{CD}, were normalized between the maximum differences of these quantities obtained for DNAs IA and IIIB. The fractional changes of the major CD subspectra, $F(C_{CD})$, are plotted versus the fractional changes in the melting free-energy, $F(-\Delta G_{Dcal})$ in Figure 7. Values for the module A (white symbols) and module B (black symbols) molecules indicate that a logarithmic relationship exists between $F(C_{CD})$ and $F(-\Delta G_{Dcal})$.

Equilibrium Binding Isotherms. The concentration-dependent melting data collected from optical melting experiments on module A and B molecules (see Figure 3) provided a means of extrapolating melting temperatures to the much lower DNA concentrations where *Bam*HI binding assays were performed. At the nanomolar concentrations used in the binding assays, the linearly extrapolated t_m values of the six molecules ranged from 35 to 50 °C, well above the temperature (18 °C) where gel shift assays were performed. Therefore, under conditions of gel shift assays, it was assumed DNA substrates resided primarily in the duplex state.

Typical results of gel shift assays in which *Bam*HI was titered against the six DNA substrates in Figure 1 are shown in the autoradiogram in Figure 8A. The particular gel depicted in Figure 8A shows results of gel shift assays of DNA IB (see Figure 1). Prior to the addition of *Bam*HI (lane 1), two bands were present, one corresponding to the duplex (D) and the other to the labeled single strand (not seen in this particular gel because it ran off the gel). Since all DNAs contained a slight excess of labeled single-stranded DNA, this band could be clearly visualized for every sample by reducing the electrophoresis run time. A third band (DE, the second band seen in Figure 8) with significantly retarded mobility emerges and becomes more prominent with increasing *Bam*HI concentrations (lanes 2−15). Concomitant with the appearance of the third band, the intensity of the duplex band (D) decreases. This third band is attributed to the complex of *Bam*HI and the duplex DNA assuming a single *Bam*HI dimer is bound at the cognate site in the center of the duplex. At even higher *Bam*HI concentrations [≥8 ×

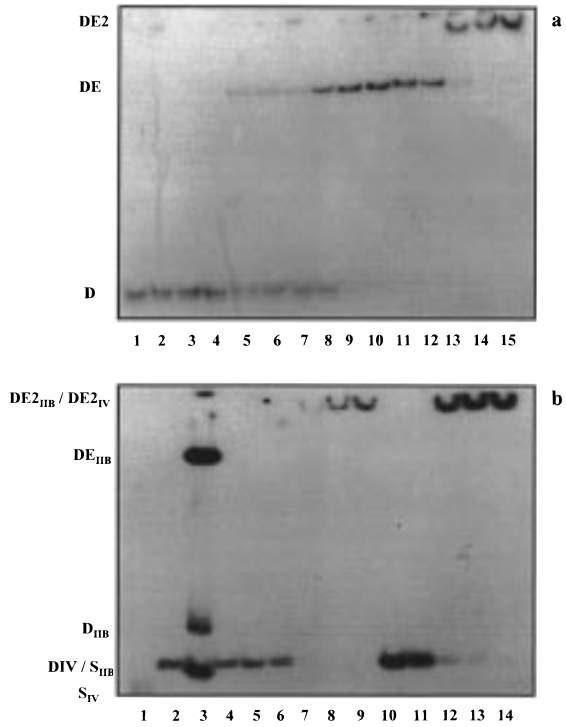

FIGURE 8: Autoradiogram for gel binding analysis. (A) Representative binding reaction and electrophoresis conditions for assaying binding to the cognate site. The duplex is DNA IA in Figure 1 at a concentration of 1×10^{-10} M: lane 1, DNA IA alone; and lanes 2−15, DNA IA titrated with 1×10^{-10}, 3×10^{-10}, 5×10^{-10}, 8×10^{-10}, 1×10^{-9}, 1.5×10^{-9}, 2×10^{-9}, 2.5×10^{-9}, 3×10^{-9}, 5×10^{-9}, 8×10^{-9}, 1×10^{-8}, 3×10^{-8}, 5×10^{-8} M *Bam*HI, respectively. Migration positions for the unbound duplex (D), the once-bound specific complex of *Bam*HI and DNA (DE), and the secondary complex (DE2) are denoted. (B) Autoradiogram of the gel retardation assay for investigating secondary mode binding using a 17-base pair deoxyoligonucleotide (DNA IV; see the text) whose sequence does not contain a cognate *Bam*HI site. Reaction and electrophoretic conditions are described in the text. DNA IV was titrated with increasing amounts of *Bam*HI: lane 1, labeled single strand of duplex IV (S_{IV}); lane 2, duplex IV alone (D_{IV}); and lane 3, 22-base pair duplex DNA IIB of Figure 1 (D_{IIB}) at a concentration of 1×10^{-10} M, an excess single strand of duplex IIB (S_{IIB}), and 3×10^{-9} M *Bam*HI. The positions of the specific (DE_{IIB}) and secondary complexes ($DE2_{IIB}$) of DNA IIB are shown. Lanes 4−9 contain 1×10^{-10} M duplex IV with increasing amounts of enzyme at 8×10^{-9} (lane 4), 1×10^{-7} (lane 5), 5×10^{-7} (lane 6), 3×10^{-6} (lane 7), 5×10^{-6} (lane 8), and 1×10^{-5} M (lane 9). Lanes 10−14 contain 1×10^{-9} M duplex IV with increasing amounts of enzyme: lane 10, 1×10^{-7} M *Bam*HI; lane 11, 5×10^{-7} M *Bam*HI; lane 12, 3×10^{-6} M *Bam*HI; lane 13, 5×10^{-6} M *Bam*HI; and lane 14, 8×10^{-6} M *Bam*HI. As indicated, the migration position of the bound complex of DNA IV ($DE2_{IV}$) is the same as for the secondary complex of DNA IIB ($DE2_{IIB}$).

10^{-7} M for DNA IB (lanes 13−15 in Figure 8)] after the unbound duplex band (D) is no longer observed (presumably after saturation of specific sites), a fourth shifted band with an even more retarded mobility (DE2) is seen. This fourth, most retarded gel band (DE2) emerges only at relatively higher *Bam*HI concentrations, above which all cognate sites are bound (evidenced by the absence of free duplex DNA). The increased intensity of this band also was titrated with increased *Bam*HI concentrations, suggesting that it corresponds to additional binding of the enzyme to these short DNAs. We term the binding attributed to this band as the "secondary" binding mode. The proposition of this secondary binding mode is not unprecedented. Others have also attributed the presence of multiple bands in gel shift assays of *Eco*RI and *Bam*HI to secondary binding (*22*−*25*). Further, in preliminary crystallization studies of *Bam*HI−DNA structures, it was shown that 12 base pairs of duplex are able to accommodate a single *Bam*HI dimer (*26*). These data and our results suggest that it is plausible that more than one *Bam*HI dimer can bind the 22-base pair duplexes.

In other experiments (not shown) in which the intensities of the fastest migrating band (excess labeled single strand) were present and quantitatively determined, they remained constant over the entire enzyme titration range. No detectable evidence for *Bam*HI binding to the component single strands of the duplexes was obtained. This was verified by independent experiments in which *Bam*HI was added in increasing amounts to solutions of individual labeled single strands, and no additional bands other than the one corresponding to single-stranded DNA alone were seen (data not shown). In other words, the intensities of the excess single strand do not effect assignment of the pixel densities of the bands corresponding to the duplex and bound complexes, or attribution to the relative concentrations of the unbound and bound DNAs. Any material in the wells was considered part of the secondary band.

To investigate whether the secondary mode binding depends on the presence of the cognate site, binding experiments were performed with a 17-base pair duplex control (DNA IV), whose sequence does not contain a GGATCC *Bam*HI cognate site. Binding assays were carried out over the same *Bam*HI concentration range as the binding experiments with the module A and B DNAs. Results of this assay are shown in Figure 8B. In the experiment that is depicted, DNA IV present at 1.0 (lanes 4−9) or 0.1 nM (lanes 10−14) was titrated with *Bam*HI over the concentration range of 10^{-7}−10^{-5} M, the protein concentration range where the secondary binding mode was observed for the module A and module B DNAs. Lane 1 contains one of the 17-base single strands of DNA IV alone (S_{IV}). Lane 2 is the 17-base pair duplex IV alone (D_{IV}). Lane 3 shows a binding reaction using 0.1 nM DNA IIB and 3×10^{-9} M *Bam*HI under conditions in which both specific and secondary mode binding are observed. There are four bands in lane 3. The lowermost band that runs slightly faster than the duplex of DNA IV is one of the 22-base single strands of duplex IIB (S_{IIB}). The next higher band in the lane corresponds to the 22-base pair duplex of DNA IIB (D_{IIB}). The next band up the gel in lane 3 corresponds to the specific complex of *Bam*HI with DNA IIB (DE_{IIB}). Finally, the fourth band seen at the top of the gel is attributed to the secondary mode complex of *Bam*HI and DNA IIB ($DE2_{IIB}$). The binding reactions for DNA IV present at two concentrations (lanes 4−9 and 10−14) with increasing *Bam*HI concentrations exhibit different behaviors. In lanes 4−6, only the duplex band for duplex IV alone (D_{IV}) is seen. At higher enzyme concentrations (lanes 7−9), the naked duplex band is not seen and a much slower band up near the well ($DE2_{IV}$), which resembles the bands in lanes 13−15 of Figure 8A, is observed. This upper band is taken to correspond to the secondary mode complex for *Bam*HI binding to DNA IV, which does not contain a cognate site. When DNA IV is present at a 10-fold higher concentration (lanes 10−14) with the same increases in the *Bam*HI concentration, the same trend is observed. At the lowest enzyme concentrations (lanes

10 and 11), only the band corresponding to duplex IV alone (D_{IV}) is observed. At higher enzyme concentrations, the secondary mode band appears along with the duplex band (lanes 12 and 13). Finally, at sufficiently high enzyme concentrations, only the secondary binding mode band is observed (lane 14).

The results of the experiments depicted in Figure 8B can be summarized as follows. First, the band corresponding to the specifically bound duplex (seen for duplex IIB in the presence of *Bam*HI, DE_{IIB}) is not seen for duplex IV. Second, the bands corresponding to the secondary mode complexes for DNAs IIB and IV migrate to similar positions (at the top of the gel), thereby providing additional evidence that at sufficiently high concentrations more than one *Bam*HI dimer is capable of binding our 22-base pair DNA oligomer substrates. If the enzyme remains a dimer at these higher concentrations, then the above result suggests binding occurs at secondary sites in addition to the cognate site. For this reason, this mode of binding is termed the secondary binding mode. Alternatively at higher concentrations, the enzyme may be forming associated structures larger than the dimer which could also lead to the greatly retarded mobility of DNAs bound in the secondary binding mode, or entrap DNAs in the wells as was observed in some cases. However, nondenaturing gel electrophoretic analysis of the enzyme alone at concentrations over and above the range used in binding experiments did not reveal any evidence for the presence of aggregates or other higher-order structures in the protein solution (not shown).

Binding curves for site-specific binding, generated from multiple gel shift assays of binding reactions between *Bam*HI and the DNAs with modules A and B, are displayed in Figure 9. Plots of the fraction of specific bound duplex DNA, F_{bS}, versus log[*Bam*HI] are shown. Symbols are averages of at least two experiments, and error bars denote standard deviations of multiple experiments. Solid curves drawn through the data were obtained by fitting each set of points with eq 5 by varying only the dissociation constant, K_D, and assuming a binding number n of 1. Both the module A and module B sequences exhibit the same trends as a function of end sequences flanking the core. That is, regardless of the module sequence, the level of binding increases with increasing %AT of the four base pairs on both ends. When all the sequences are evaluated together in terms of their overall %GC, they range from 18.2% for molecule IA to 81.8% for molecule IIIB. Comparison of the binding curves for these DNAs in Figure 9 (○ and ▲) reveals that DNAs with the widest difference in %GC in their sequences flanking the central binding site exhibit the greatest difference in binding levels. Comparison of the binding curves obtained for sequences IIA (△) and IIB (▲), which have a difference in %GC of 36.4 vs 63.6%, indicates these binding curves are closer (revealing smaller differences in observed binding levels) than for sequences IA and IIIB. This trend of convergence of the level of *Bam*HI binding with decreased differences in %GC of the flanking sequences is also seen in results for DNAs IIIA (□) and IB (●). The sequences of these duplexes have the same %GC (54.5%) and differ only in the relative order or context of the A-T and G-C type base pairs that flank the common eight-base pair core. Values of the specific dissociation constant K_{DS} are summarized in Table 3. These values, assumed to correspond to dissociation

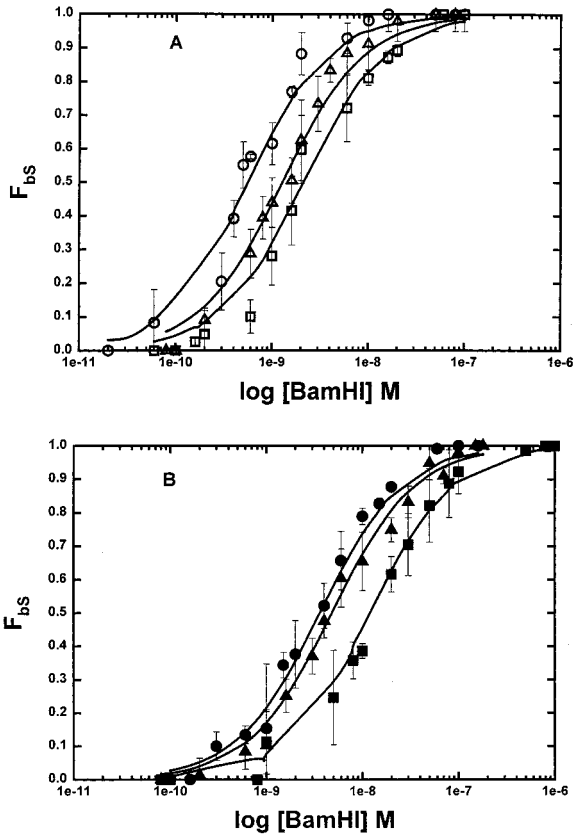

FIGURE 9: Binding isotherms for the specific binding mode. The fraction of DNA bound at the cognate site by one *Bam*HI dimer determined at 18 °C (F_{bS}) is plotted vs log[*Bam*HI]. These data were collected from gel retardation experiments carried out in the absence of Mg^{2+}. The duplex concentration was 1×10^{-10} M. (A) The binding isotherms for module A molecules IA (○), IIA (△), and IIIA (□). (B) The binding isotherms for module B molecules IB (●), IIB (▲), and IIIB (■). Curves were fit with eq 5 as described in the text and provided evaluations of the specific mode binding parameter, K_{DS}. These values are summarized in Table 3. Error bars represent the average reproducibility for at least three experiments.

Table 3

DNA	K_{DS} ($\times 10^{-9}$ M)	K_{DN} ($\times 10^{-9}$ M)	ΔG_B(20 °C) (kcal/mol)
IA	0.5 ± 0.1	6.1 ± 0.6	−12.5 ± 0.1
IIA	1.2 ± 0.2	9.3 ± 1.0	−12.0 ± 0.1
IIIA	2.1 ± 0.2	31.0 ± 5.0	−11.6 ± 0.1
IB	3.6 ± 0.4	61.0 ± 7.0	−11.3 ± 0.1
IIB	5.0 ± 0.5	27.0 ± 3.0	−11.1 ± 0.1
IIIB	12.0 ± 1.3	253.0 ± 28.0	−10.6 ± 0.1

of one *Bam*HI dimer ($n = 1$) per duplex, range from 0.5 to 12.0×10^{-9} M (factors of 2−25) depending upon the %GC of the sequence flanking the *Bam*HI cognate site in these DNA oligomers.

Results analogous to those shown in Figure 9 for the specific binding mode are displayed in Figure 10 for the secondary binding mode of *Bam*HI. The fraction of bound DNA in the secondary mode, F_{bN}, is plotted versus log[*Bam*HI]. Again, symbols are the average of at least two experiments and error bars are the experimental standard deviations for multiple experiments. Solid curves were obtained by fitting the data using eq 5, varying K_D ($=K_{DN}$) as the fitting parameter, assuming two binding sites ($n = 2$). Binding curves for the module A molecules are shown in Figure 10A, and those for the module B molecules are

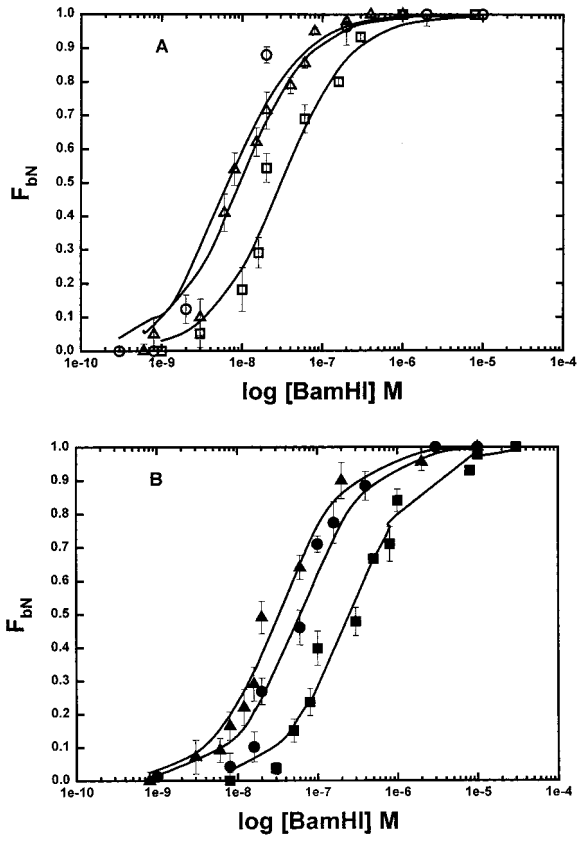

FIGURE 10: Binding isotherms for the secondary binding mode. The fraction of DNA bound in the secondary mode, F_{bN}, determined at 18 °C plotted vs log[*Bam*HI]. Fractions of slowest migrating bands observed for all DNA substrates represent presumably more than one *Bam*HI dimer bound to the same DNA substrate. Symbols and binding conditions are the same as those described in the legend of Figure 9. From theses curves, the secondary mode binding parameter, K_{DN}, was evaluated for each deoxyoligonucleotide substrate. The order of modulation for secondary mode binding follows that of specific binding, suggesting that secondary mode binding by *Bam*HI is also sequence-dependent. Values of K_{DN} are summarized in Table 3.

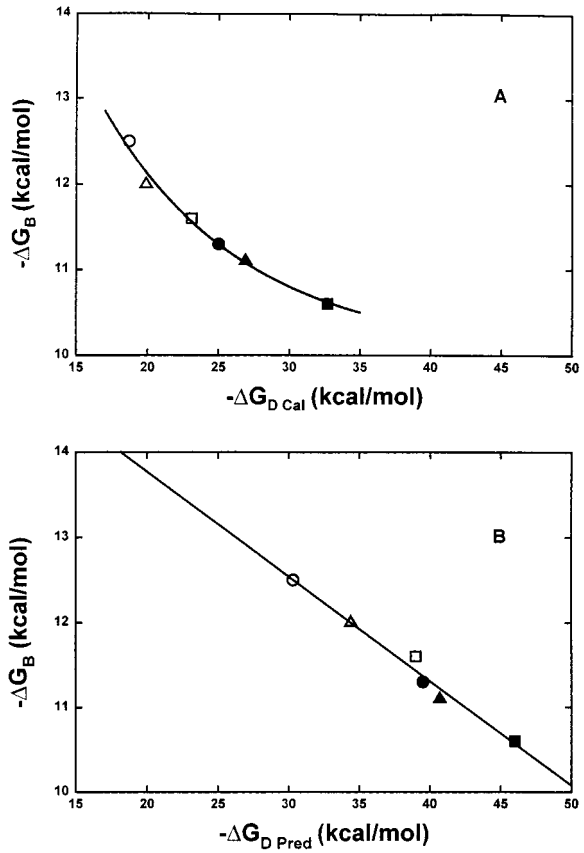

FIGURE 11: Relationships between the free energy of enzyme binding and the free energy of flanking sequence stability. (A) Values of the binding free energy, ΔG_B, evaluated from the binding constant, K_{DS} [$\Delta G_B = -RT \ln(K_{DS})$], as summarized in Table 3, are plotted vs the free energies of duplex melting, ΔG_{Dcal}, as summarized in Table 1. Data for module A molecules IA (○), IIA (△), and IIIA (□). Data for module B molecules IB (●), IIB (▲), and IIIB (■). (B) Plot of ΔG_B for the module A and module B molecules vs their melting free energies, ΔG_{pred}, predicted from published nearest-neighbor stability parameters. Symbols are the same as for panel A. Free energy values are at 20 °C.

shown in Figure 10B. Interestingly, the same sequence-dependent trends that were seen for the specific binding mode in Figure 9 are found in Figure 10 for secondary mode binding. The extent of binding decreases with increases in the %GC of the DNAs. Dissociation constants for the secondary mode, K_{DN}, evaluated from the data presented in Figure 10 are summarized in Table 3. Data presented in Figures 9 and 10 and Table 3 reveal a common trend. The greater the differences in the %GC of sequences flanking the common core sequence, the greater the observed differences in the level of equilibrium binding of the *Bam*HI restriction endonuclease.

As Table 3 clearly shows, the dissociation constants estimated from gel shift experiments for the secondary binding mode of the module A and B DNAs vary with flanking sequences. For DNA IV, the value of the observed dissociation constant of the secondary binding mode is ~400 $\times$ 10^{-9} M, a factor of 2−200 times greater than the dissociation constants found for the secondary binding modes of the module A and B DNAs given in Table 3. We were able to construct separately and fit the specific and secondary binding isotherms because the naked duplex and the secondary bound complex were never observed in the same gel lane at the same protein concentration. Further, in no case was the ratio of the fraction of DNA bound in the specific mode to that bound in the secondary mode F_{bS}/F_{bN} less than 4; $F_{bN} \leq 0.2$ until $F_{bS} \geq 0.8$. That is, the isotherms were sufficiently separated over the protein concentration range that was examined to allow significant fits with the model.

Relationships between Binding and Melting. To compare enzyme binding and duplex stability, the free energy of *Bam*HI binding determined from the dissociation constants summarized in Table 3, $-\Delta G_B$, is plotted versus the melting free energy determined by DSC, $-\Delta G_{Dcal}$, in Figure 11A. White symbols represent the data for the module A molecules, and black symbols represent the data for the module B molecules. In this plot, $-\Delta G_B$ varies in a nonlinear manner with $-\Delta G_{Dcal}$. The line shown is a decaying single-exponential fit.

We also wished to see how the predicted free energy of the sequences, $-\Delta G_{Dpred}$, which neglects the contribution of duplex initiation, compared with $-\Delta G_B$. This comparison is depicted in the plot of ΔG_B versus $-\Delta G_{Dpred}$ shown in Figure 11B, which is linear. It was shown in Figure 5 that the helix initiation free energy, ΔG_{int}, is sequence-dependent, suggesting the presence of additional types of sequence-dependent effects, other than base pair hydrogen bonding and nearest-neighbor stacking, that may contribute to the

stability (and melting) of these DNAs. Effects of duplex initiation would be expected to be most significant in the melting of short DNA oligomers. As base pair length increases, the influence of ΔG_{int} becomes increasingly small. The dominant free energy component for the long DNA limit then becomes that predicted from the n-n sequence-dependent parameters, ΔG_{pred}. Thus, the observed nonlinear relationship between $-\Delta G_B$ and $-\Delta G_{Dcal}$ may be due to the sequence dependence of helix initiation for these short deoxyoligonucleotides. The linearity between $-\Delta G_B$ and $-\Delta G_{Dpred}$, which does not include helix initiation, suggests the observed dependence of ΔG_B on ΔG_{Dpred} could persist to greater distances for longer flanking sequences. Studies aimed at testing this conjecture are currently underway (P. M. Vallone and A. S. Benight, manuscript in preparation).

All molecules contain the *Bam*HI enzyme cognate binding site. In our assays when probed with *Bam*HI, the flanking sequences significantly influence *Bam*HI binding. The observed effect of flanking sequence occurs at least four base pairs away from the cognate binding site. Since the two module sequences exhibit the same trends, increasing %GC results in relatively poorer binding.

The free energy of duplex initiation or nucleation, ΔG_{int}, depends on both the module and end sequences. Greater differences in ΔG_{int} values were found for different molecules depending on the module sequences. There are two obvious potential reasons for this observation. First, perhaps the predicted free energies obtained from n-n values are incorrect for these molecules because of the presence of significant interactions longer than nearest-neighbor. Second, from comparisons of results from optical and calorimetric melting experiments, an assessment of the two-state versus non-two-state melting behavior of the molecules was obtained. This analysis indicated that four of the six DNAs exhibit non-two-state melting transitions, indicating that even DNA oligomers of the same short duplex length (22 base pairs) can deviate from two-state melting depending on the sequence. We have attributed at least part of this effect to the sequence dependence of duplex initiation and end effects.

Flanking Sequence and Site-Specific Binding at the Cognate Site. It has been observed for some time that activities (binding and cleavage) of DNA restriction enzymes at specific sequence "cognate" sites on duplex DNA can be significantly influenced by the context of surrounding flanking sequences. Such context effects have been reported for DNA strand cleavage by *Eco*RI, *Hin*fI, *Pst*I, *Fnu*DII, *Hae*III, *Hha*I, and *Msp*I enzymes at their cognate sites (*27*−*30*). Although the origins of this apparently generic sequence-dependent effect have not been clearly defined, studies with DNA oligomers having a single cognate site provide a means with which to dissect and evaluate subtle effects of different flanking sequences on enzyme binding (*27*, *31*, *32*).

In earlier studies, we measured rates of first-strand cleavage by the *Alu*I restriction enzyme at its recognition sequence 5′-A-G-C-T-3′, imbedded in short duplex DNA oligomers flanked by slightly different sequence environments with variable sequence-dependent thermodynamic stabilities (*31*). Binding-limited cleavage rates were found to be inversely proportional to the sequence-dependent stability of the duplex substrates. Consistent with the results that are reported here, it was gleaned that restriction enzyme reactivity at a specific sequence can be modulated in an incremental manner by the differential melting stability of sequences flanking the site.

In the studies presented here, we developed and employed an assay to study the effects of different flanking sequence contexts on equilibrium binding of the *Bam*HI endonuclease. *Bam*HI binding was investigated for six different 22-base pair deoxyoligonucleotide DNAs each containing the *Bam*HI cognate site in the center. All 22-base pair substrates contain the same eight base pairs in the center that includes the six-base pair cognate site where the ligand (in this case restriction enzyme *Bam*HI) binds. If the cognate binding site is not significantly perturbed by the flanking sequence, by virtue of the assay system design, all interactions between the enzyme and DNA (at the cognate site) would be expected to be the same for these molecules, and their binding constants would be expected to be equivalent. This expectation is not met, since binding is in fact influenced by flanking sequences. The influence of flanking sequence on binding allows us to infer that the hydrogen bonds, van der Waals, electrostatic, and solvent interactions, and other important direct and indirect structurally related contacts between the enzyme and DNA required for formation of a sequence-specific complex must vary for different flanking sequences. Observed differences in binding behavior for DNA substrates with the same length and core sequence, but different flanking sequences, support this proposition.

Koudelka et al. (*33*) found the sequence of the noncontacted four base pairs in the central sequence between the half-sites bound by the lambda 434 repressor strongly influenced protein binding. Changing the sequence from all A-T to all G-C resulted in a 50-fold decrease in binding affinity. To explain these observations, it was proposed that the twisting rigidity of A-T base pairs is lower than that of G-C base pairs, and thereby the more torsionally flexible A-T sequence facilitates binding and equilibrium complex formation. However, Fujimoto and Schurr (*34*) convincingly argued that the torsional (or twisting) rigidity of DNA is virtually independent of sequence and that instead the observed sequence effect must be attributed to differences in equilibrium structure and electrostatic and/or bending rigidity. Since we find similar correlations between the A-T content of flanking sequences and binding affinity, perhaps our results can be explained similarly. That is (as suggested above), different flanking sequences may present sequence-dependent variations in equilibrium structure and electrostatics that in turn influence binding at the cognate site.

Melting Stability and Enzyme Binding. If the DNA duplex substrates are bound at temperatures sufficiently below their melting temperatures such that the duplex state prevails (presumably the case for our experiments) and the duplex does not have to be melted to be bound by the enzyme, or is not melted by the enzyme in the process of complex formation, then it would not be expected a priori that the stability of the flanking sequence would affect site-specific binding at the same cognate site. At present, it is not clear from a mechanistic point of view how the melting potential of a duplex substrate can effect enzyme binding to that duplex. If melting stability differences correlate to differences in structures adopted by flanking sequences, or structural differences imparted by flanking sequences on the cognate site, these would be expected in turn to differentially affect enzyme binding. In the case of this study, melting stability

was found to be a direct predictor of how variations in flanking sequence affect enzyme binding.

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. Portions of this work appeared in the Ph.D. Thesis of Peter V. Riccelli, Department of Chemistry, University of Illinois, Chicago, December, 1997.

REFERENCES

1. Owczarzy, R., Vallone, P. M., Gallo, F. J., Paner, T. M., Lane, M. J., and Benight, A. S. (1997) *Biopolymers 44*, 217−239.
2. Allawi, H. T., and SantaLucia, J., Jr. (1997) *Biochemistry 36*, 10581−10594.
3. Breslauer, K. J., Frank, R., Blocker, H., and Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746−3750.
4. Caruthers, M. H. (1982) *Chemical and Enzymatic Synthesis of Gene Fragments*, Verlag-Chemie, Weinheim, Germany.
5. Cantor, C. R., Warshaw, M. M., and Shapiro, H. (1970) *Biopolymers 9*, 1059−1077.
6. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1992) *Molecular Cloning*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
7. 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.
8. Amaratunga, M., Pancoska, P., Paner, T. M., and Benight, A. S. (1990) *Nucleic Acids Res. 18*, 577−582.
9. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1988) *Numerical Recipes*, Cambridge University Press, New York.
10. Sheardy, R. D., Suh, D., Kurzinsky, R., Doktycz, M. J., Benight, A. S., and Chaires, J. B. (1993) *J. Mol. Biol. 231*, 475−488.
11. Nardone, G., and Chirikjian, J. G. (1987) *Gene Amplification and Analysis*, Elsevier Science, New York.
12. Aida, M. (1988) *J. Theor. Biol. 130*, 327−335.
13. Delcourt, S. G., and Blake, R. D. (1991) *J. Biol. Chem. 266*, 15160−15169.
14. Doktycz, M. J., Goldstein, R. F., Paner, T. M., Gallo, F. J., and Benight, A. S. (1992) *Biopolymers 32*, 849−864.
15. Gotoh, O., and Tagashira, Y. (1981) *Biopolymers 20*, 1033−1042.
16. McCampbell, C. R., Wartell, R. M., and Plaskon, R. R. (1989) *Biopolymers 28*, 1745−1758.
17. SantaLucia, J., Allawi, H. T., and Seneviratne, P. A. (1996) *Biochemistry 35*, 3555−3562.
18. Sugimoto, N., Nakano, S., Yoneyama, M., and Honda, K. (1996) *Nucleic Acids Res. 24*, 4501−4505.
19. Vologodskii, A. V., Amirikyan, B. R., Lyubchenko, Y. L., and Frank-Kamenetskii, M. D. (1984) *J. Biomol. Struct. Dyn. 2*, 131−148.
20. Wartell, R. M., and Benight, A. S. (1985) *Phys. Rep. 126*, 67−107.
21. SantaLucia, J. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 1460−1465.
22. George, J., Nardone, G., and Chirikjian, J. G. (1985) *J. Biol. Chem. 260*, 14387−14392.
23. Grabowski, G., Jeltsch, A., Wolfes, H., Maass, G., and Alves, J. (1995) *Gene 157*, 113−118.
24. Xu, S. Y., and Schildkraut, I. (1991) *J. Bacteriol. 173*, 5030−5035.
25. Xu, S. Y., and Schildkraut, I. (1991) *J. Biol. Chem. 266*, 4425−4429.
26. Strzelecka, T., Newman, M., Dorner, L. F., Knott, R., Schildkraut, I., and Aggarwal, A. K. (1994) *J. Mol. Biol. 239*, 430−432.
27. Alves, J., Pingoud, A., Haupt, W., Langowski, J., Peters, F., Maass, G., and Wolff, C. (1984) *Eur. J. Biochem. 140*, 83−92.
28. Armstrong, K., and Bauer, W. K. (1982) *Nucleic Acids Res. 10*, 993−1007.
29. Drew, H. R., and Travers, A. A. (1985) *Nucleic Acids Res. 13*, 4445−4467.
30. Goldstein, K., Thomas, M., and Davis, R. (1975) *Virology 66*, 420−427.
31. 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.
32. Taylor, J. D., and Halford, S. E. (1992) *Biochemistry 31*, 90−97.
33. Koudelka, G. B., Harbury, P., Harrison, S. C., and Ptashne, M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4633−4637.
34. Fujimoto, B. S., and Schurr, J. M. (1990) *Nature 344*, 175−178.

BI9904407