An FTIR Investigation of Flanking Sequence Effects on the Structure and Flexibility of DNA Binding Sites[†]

Talia R. Kahn,[‡] Kimberly K. Fong, Brian Jordan, Janista C. Lek, Rachel Levitan,[§] Patrick S. Mitchell,[||] Corrina Wood,[⊥] and Mary E. Hatcher*

Joint Science Department, Claremont McKenna, Pitzer, and Scripps Colleges, Claremont, California 91711

Received August 12, 2008; Revised Manuscript Received November 15, 2008

ABSTRACT: Fourier transform infrared (FTIR) spectroscopy and a library of FTIR marker bands have been used to examine the structure and relative flexibilities conferred by different flanking sequences on the *Eco*RI binding site. This approach allowed us to examine unique peaks and subtle changes in the spectra of d(AAAGAATTCTTT)$_2$, d(TTCGAATTCGAA)$_2$, and d(CGCGAATTCGCG)$_2$ and thereby identify local changes in base pairing, base stacking, backbone conformation, glycosidic bond rotation, and sugar puckering in the studied sequences. The changes in flanking sequences induce differences in the sugar puckers, glycosidic bond rotation, and backbone conformations. Varying levels of local flexibility are observed within the sequences in agreement with previous biological activity assays. The results also provide supporting evidence for the presence of a splay in the G_4-C_9 base pair of the *Eco*RI binding site and a potential pocket of flexibility at the G_4 cleavage site that have been proposed in the literature. In sum, we have demonstrated that FTIR is a powerful methodology for studying the effect of flanking sequences on DNA structure and flexibility, for it can provide information about the local structure of the nucleic acid and the overall relative flexibilities conferred by different flanking sequences.

During both replication and transcription, amino acids of the protein factor involved recognize and interact electrostatically with the base pairs of the specific DNA sequence containing the recognition site (direct readout) with remarkable specificity. Upon binding of a protein in this region, a bending of the DNA occurs which enhances the stability of the complex by increasing the number of favorable contacts made between the DNA and the protein (*1, 2*). The four classifications of these interactions are (i) van der Waals interactions between aliphatic amino acids and the C5-methyl groups of thymines or the C5-hydrogen groups of cytosines, (ii) hydrogen bonds between the nucleic acids and the amide and carbonyl groups of the peptide backbone, (iii) hydrogen-bonding bridges constructed with water molecules between amino acids and base pairs (*3*−*5*), and (iv) interlacing of amino acid side chains between two bases (*6*). Although these are the major interactions, direct contacts at the recognition site alone do not account for the conformation of the DNA−protein complex; base pairs on either side of the recognition site, referred to as flanking sequences, are also thought to play a role in the conformation of the DNA−protein complex (indirect readout) (*1, 7, 8*).

In most cases, it is necessary for the protein to distort the DNA in order to achieve the transition state. To compensate for the energetic costs of deformation, the DNA−protein complex may make any modification that improves the binding free energy such as releasing water from nonpolar surfaces and/or increasing the number of direct contacts made between the DNA and protein (*9*). DNA that requires less deformation to reach the conformation of the transition state will have a lower free energy of association (*10*). Similarly, sequences that have an inherently greater distortability will also have a lower binding free energy. Studies examining the effect of flanking sequences on DNA−protein binding have indicated that flanking sequences are involved in dynamics and confer flexibility or rigidity to a sequence (*1, 7, 8, 11*−*14*). Since flexible DNA has a greater distortability, it is able to compensate more effectively and thereby more efficiently recognize and bind proteins and drugs than a similar, but more rigid DNA sequence (*15, 16*).

Various methods have been commonly used to examine the effect of flanking sequences on DNA including conformational free energy calculations (*8*), ^{1}H NMR and ^{31}P NMR (*11, 17, 18*), crystal structures (*3, 36*), restriction enzyme digests (*12, 13, 19*), molecular dynamics simulations (*20*), UV melting (*14, 19*), sequence-dependent anisotropic bendability (SDAB) (*21*), and CD (*14, 19, 22*). These techniques have revealed valuable information about the features in DNA that change when flanking sequences are modified including changes in the unwinding and overwinding of the double helix, formation of hydrogen bonds, twist and glycosidic angles, helical parameters, flexibility, and restriction endonuclease cutting rates (*3, 11, 12, 20*). Although these methods are highly advanced, there is some

[†] This work was supported by a grant from the NIH (1R15CA117891).

* To whom correspondence should be addressed. Phone: (909) 607-1586. Fax: (909) 621-8588. E-mail: mhatcher@jsd.claremont.edu.

[‡] Current address: University of Washington School of Medicine, Seattle, WA.

[§] Current address: Tulane University School of Medicine, New Orleans, LA.

[||] Current address: Fred Hutchinson Cancer Research Center, Seattle, WA.

[⊥] Current address: New York University School of Medicine, New York, NY.

10.1021/bi8015235 CCC: $40.75
Published on Web 01/23/2009

disparity with regard to which sequences are most flexible and most rigid (*11, 12, 21*).

In principle, Fourier transform infrared (FTIR) spectroscopy can detect local conformational changes with high sensitivity (*16, 23, 24*). For example, FTIR has been used to study DNA structural changes caused by CpG methylation (*16*) and the structural effects of G-tract length (*22*). Since this technique is nondestructive and provides a "snapshot" of the nucleic acid's structure, we have chosen to apply FTIR along with a library of IR bands compiled by Banyay et al. (*23*) to interpret major features and changes in nucleic acid structure, conformations, and conformational substates as a result of different flanking sequences.

MATERIALS AND METHODS

Sample Preparation. The DNA dodecamers studied were obtained from Alpha DNA (Montreal, Quebec, Canada). Samples were desalted by centrifugation in Y300 Centricons (Millipore, Billerica, MA). After desalting, samples were lyophilized overnight and stored at 4 °C until analysis.

Fourier Transform Infrared (FTIR) Experiments. Samples for FTIR analysis were prepared by diluting to a final concentration of 3.0 mM in 100 mM sodium phosphate buffer (pH 7.4). Freshly made samples were used for all FTIR measurements. FTIR spectra were recorded at room temperature on a ZnSe crystal using a Thermo Nicolet Avatar 370 ATR-FTIR spectrometer (Thermo Scientific, Waltham, MA). A background (atmospheric plus pure buffer) was collected after which DNA samples were run. For each DNA sample, FTIR spectra were taken in triplicate to ensure reproducibility. Experiments were run with a resolution of 4 cm^{-1} and 512 scans. Using EZ-OMNIC software (Thermo Scientific), all DNA spectra underwent baseline correction and 15 point smoothing before peaks were assigned. Derivative spectra were 25 point smoothed. Final spectra were normalized with respect to the symmetric PO_2^- stretching mode appearing at 1090−1080 cm^{-1} (*25, 26*). Spectral assignments were made using the Banyay et al. library (*23*). This library provides wavenumber ranges of ∼5−10 cm^{-1} for many assignments so differences less than 10 cm^{-1} may not be significant.

^{31}P NMR Experiments. DNA samples were obtained from Alpha DNA. These were salted with 100 mM phosphate buffer at pH = 7.4 and then lyophilized. The samples were rehydrated in 98% D_2O, relyophilized, and finally dissolved in 99.9% D_2O, immediately before use.

Two-dimensional ^{1}H nuclear Overhauser effect (NOESY) experiments were performed with mixing times of 500 ms and a gradient pulse applied during the mixing time. A total of 4096 data points were collected in the t_2 dimension and 512 increments in t_1. The matrix was zero-filled to 4096 × 1024 points before double Fourier transformation. Phase-sensitive ^{31}P−^{1}H HSQC were recorded using Echo/Antiecho-TPPI gradient selection (*27*−*29*) and decoupling during acquisition. All HSQC were performed with 2048 points in the t_2 dimension and 256 t_1 increments using a relaxation delay of 4 s. The chemical shift of H_2O and the center of the DNA phosphorus resonances were set as the center of HSQC and NOESY spectra. A spectral width of 9 ppm was used for all 1D experiments and in both dimensions of the 2D experiments.

Table 1: Sequences of DNA Samples Studied[a]

DNA sample	sequence
*Eco*RI AAA/TTT	d(AAA**GAATTC**TTT)$_2$
*Eco*RI TTC/GAA	d(TTC**GAATTC**GAA)$_2$
*Eco*RI CGC/GCG	d(CGC**GAATTC**GCG)$_2$

[a] The bold lettering indicates the *Eco*RI binding site.

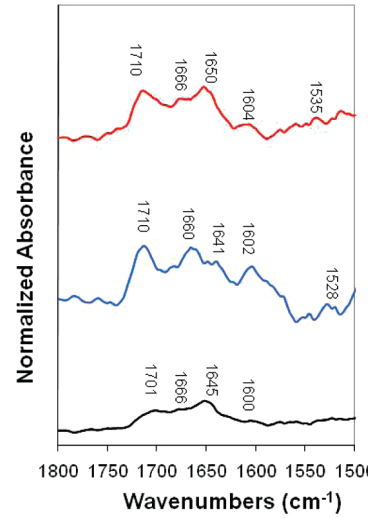

FIGURE 1: Spectra for the 1800−1500 cm^{-1} in-plane base vibrations: *Eco*RI AAA/TTT (top), *Eco*RI TTC/GAA (middle), and *Eco*RI CGC/GCG (bottom).

H2′ and H2″ are strongly coupled with H1′, H3′, H5′ (sugar protons) and to the base protons, enabling the assignment of protons on the NOESY using a sequential walk (*30*). The strong correlations between ^{31}P and H3′ and H5′/H4′ then enabled us to assign the HSQC using the proton assignments determined from the NOESY.

Percent BII was determined as a function of the ^{31}P chemical shift using a linear relationship between the observed fast-exchange chemical shift and the percent BII conformation (*31*):

$$\%\ \mathrm{BII} = \frac{1}{2.368 - 0.005110T}\delta P(\mathrm{ppm}) + \frac{1.345 - 0.002031T}{2.368 - 0.005110T}$$

RESULTS

We studied DNA sequences containing the *Eco*RI binding site as this binding site has been extensively studied by a number of techniques. In particular, we used the well-characterized Dickerson sequence (*32*) as a control so we can compare our results with the wealth of structural and dynamical data available in the literature on this sequence. Flanking sequences were chosen based on previous research on the effect of flanking sequences on DNA structure, dynamics, and protein−DNA interactions (*11, 12, 18, 26*). Table 1 lists the *Eco*RI sequences that were examined using FTIR. The spectral peaks were assigned using the Banyay Library of DNA infrared bands (*23*).

Effect of Flanking Sequences on the Base Region (1800−1500 cm^{-1}). Figure 1 shows the 1800−1500 cm^{-1} region for *Eco*RI binding site containing sequences. This region reports on changes in the bases of nucleic acids (*23*). The vibrations of base-paired pyrimidines occur between 1670−1630 cm^{-1}. The *Eco*RI TTC/GAA sequence has a broad range of peaks within this region including distinct

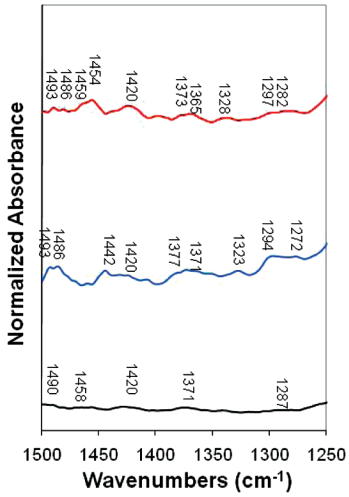

FIGURE 2: Spectra for the 1500−1250 cm^{-1} base−sugar vibrations: *Eco*RI AAA/TTT (top), *Eco*RI TTC/GAA (middle), and *Eco*RI CGC/GCG (bottom).

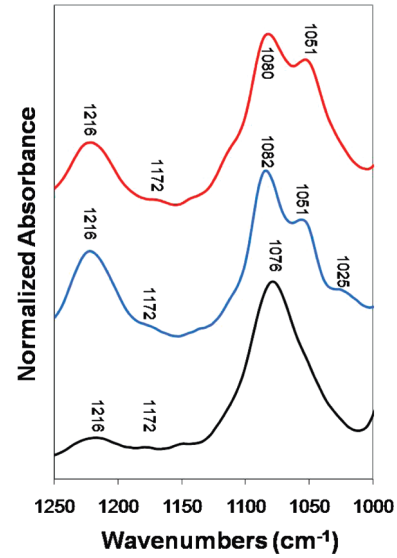

FIGURE 3: Spectra for the 1250−1000 cm^{-1} sugar−phosphate vibrations: *Eco*RI AAA/TTT (top), *Eco*RI TTC/GAA (middle), and *Eco*RI CGC/GCG (bottom).

peaks at 1660 and 1641 cm^{-1}. The peak at 1660 cm^{-1} falls into the range for single-stranded guanines (1673−1660 cm^{-1}). *Eco*RI AAA/TTT has similar breadth in this region but differing intensities. *Eco*RI CGC/GCG has a broad peak with a clear resonance at 1645 cm^{-1}. The strong resonance at 1645 cm^{-1} is indicative of single-stranded cytosines. The breadth and lack of resolution in this region for the *Eco*RI CGC/GCG sequence may be the result of the large number of CG base pairs in this sequence. A peak between 1604 and 1600 cm^{-1} was present in all samples (although only weakly in *Eco*RI CGC/GCG) and is also a marker for single-stranded guanines. The vibrations of base-paired cytosines occur at 1528 cm^{-1} in *Eco*RI TTC/GAA and shift upfield to 1535 cm^{-1} in *Eco*RI AAA/TTT. There is a low intensity, broad resonance in this region for *Eco*RI CGC, again likely due to the increased number of CG pairs. An intense peak at 1710 cm^{-1} is evident *Eco*RI AAA/TTT and *Eco*RI TTC/GAA but only weakly present in *Eco*RI CGC/GCG. The Banyay Library attributes this peak to guanines involved in Hoogsteen binding to a third strand, but this is unlikely in the short DNA sequences at low concentrations studied here. Therefore, these peaks are more likely the carbonyl stretches of bases in A-form (*33*).

Effect of Flanking Sequences on the Base−Sugar Region (1500−1250 cm^{-1}). The spectrum for the 1500−1250 cm^{-1} region demonstrating the effect of flanking sequence changes on the glycosidic bond rotation, the backbone conformation, and the sugar puckering modes of nucleic acids is illustrated in Figure 2. In *Eco*RI CGC/GCG, sugar ring vibrations of purines occur as a weak, broad peak centered at 1490 cm^{-1}. This becomes two peaks at 1493 and 1486 cm^{-1} in *Eco*RI TTC/GAA and weakly in *Eco*RI AAA/TTT. Changes in the position of this band reflect interactions at N7 sites of purines (*23*). Similarly, the broad peak due to the vibrations of adenines in A-/B-form conformation at 1458 cm^{-1} in *Eco*RI CGC/GCG splits into two distinct peaks at 1459 and 1454 cm^{-1} in *Eco*RI AAA/TTT. This peak shifts to 1442 cm^{-1} in *Eco*RI TTC/GAA. In *Eco*RI AAA/TTT there is a peak at 1420 cm^{-1}, which corresponds to S-type sugars in B-form helices; this marker is weakly present in *Eco*RI TTC/GAA and *Eco*RI CGC/GCG. A broad peak encompassing the Z-form marker at 1434−1440 cm^{-1} is present in *Eco*RI TTC/GAA. Peaks at 1373 and 1365 cm^{-1} for purines in *anti* form are present in *Eco*RI AAA/TTT, and a broad resonance encompassing this region is present in *Eco*RI TTC/GAA. A weak, broad resonance at 1371 cm^{-1} is present in *Eco*RI CGC/GCG. This shift suggests modified glycosidic torsion angles (*16*). In *Eco*RI AAA/TTT and *Eco*RI TTC/GAA, a marker for thymines with S-type sugars and a glycosyl torsional angle χ in *anti* conformation is evident at 1328 and 1323 cm^{-1}, respectively, but absent in *Eco*RI CGC/GCG. A marker corresponding to C4NH2 vibrations of a cytosine occurs in all samples between 1297 and 1287 cm^{-1}. The CN3H bends of thymines with S-type sugars and *anti* conformations occur in *Eco*RI AAA/TTT at 1282 cm^{-1}; however, this marker appears only very weakly in *Eco*RI TTC/GAA and not at all in *Eco*RI CGC/GCG.

Effect of Flanking Sequences on the Sugar−Phosphate Region (1250−1000 cm^{-1}). Figure 3 illustrates marker bands between 1250 and 1000 cm^{-1} that identify changes in the backbone conformation. All *Eco*RI sequences exhibited a marker at 1216 cm^{-1}, which corresponds to B-form DNA. All three samples also demonstrate symmetric PO^{2-} vibrations that are insensitive to the B-to-A transition between 1082 and 1076 cm^{-1}. The shoulder to this symmetric PO^{2-} peak at 1051 cm^{-1} is evident in *Eco*RI AAA/TTT and *Eco*RI TTC/GAA but is subsumed into a broad resonance in *Eco*RI CGC/GCG. This vibration is attributed to the CO stretch of the backbone and is strongly enhanced in Z-form DNA (*25, 26*). In *Eco*RI TTC/GAA, a marker for furanose vibrations in Z-DNA is evident at 1025 cm^{-1}. Finally, all three sequences have a weak marker at 1172 cm^{-1}, an A-form marker due to C3′ *endo*/*anti* puckering.

Effect of Flanking Sequences on the Sugar Region (1000−800 cm^{-1}). The effects of sugar conformation on *Eco*RI spectra provided in the 1000−800 cm^{-1} region are depicted in Figure 4. Careful examination of this region, particularly between 900 and 800 cm^{-1}, was required because important structural differences often manifested as minor spectral changes.

To aid in characterizing peaks contributing to the broad resonances observed, second derivative spectra for this region

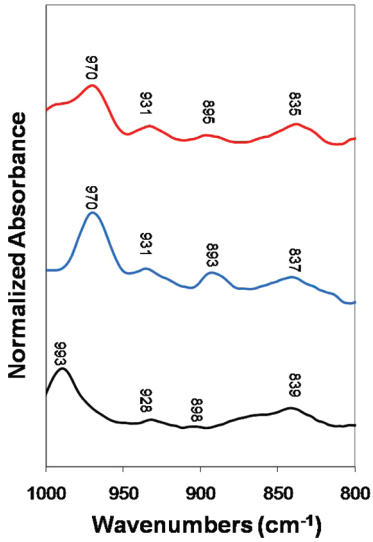

FIGURE 4: Spectra for the 1000−800 cm^{-1} sugar vibrations: *Eco*RI AAA/TTT (top), *Eco*RI TTC/GAA (middle), and *Eco*RI CGC/GCG (bottom).

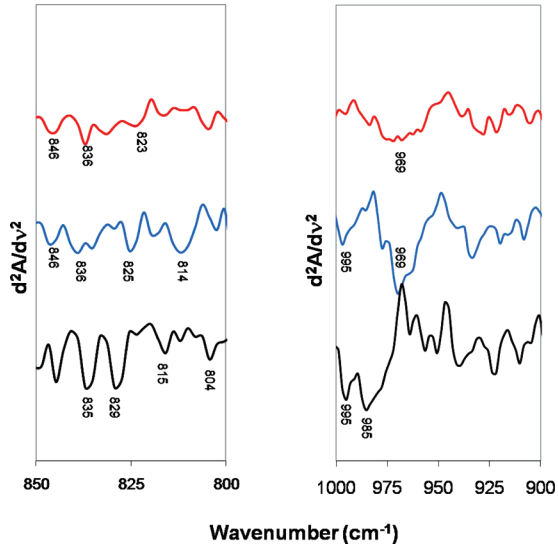

FIGURE 5: The second derivative spectra of the FTIR region reporting on the backbone and sugar conformations: *Eco*RI AAA/TTT (top), *Eco*RI TTC/GAA (middle), and *Eco*RI CGC/GCG (bottom).

are provided in Figure 5 (*16*). A significant peak at 970 cm^{-1} correlating to the CC stretch of the B-form backbones is present in *Eco*RI AAA/TTT and *Eco*RI TTC/GAA. *Eco*RI CGC/GCG and, to a much less intensity, *Eco*RI AAA/TTT demonstrate a resonance at 993 cm^{-1}. The derivative spectra show that *Eco*RI AAA/TTT has breadth, but all of its intensity is localized about the 970 cm^{-1} peak. *Eco*RI TTC/GAA has a peak at 995 cm^{-1} and a broader one at 969 cm^{-1} with additional resolution.

*Eco*RI CGC/GCG has a narrow peak at 995 cm^{-1} and a broad peak at 985 cm^{-1}. This suggests a large variety of backbone conformations in these samples and some dynamic averaging between them. The backbone dynamics of the *Eco*RI CGC/GCG sequence have been previously studied and demonstrate facile interconversion of the BI and BII states in the C3pG4 and C9pG10 steps (*31, 34*). Figure 6 shows the percent BII as a function of backbone step for each sequence as determined by ^{31}P NMR. This plot demonstrates that the *Eco*RI AAA/TTT samples mainly BI

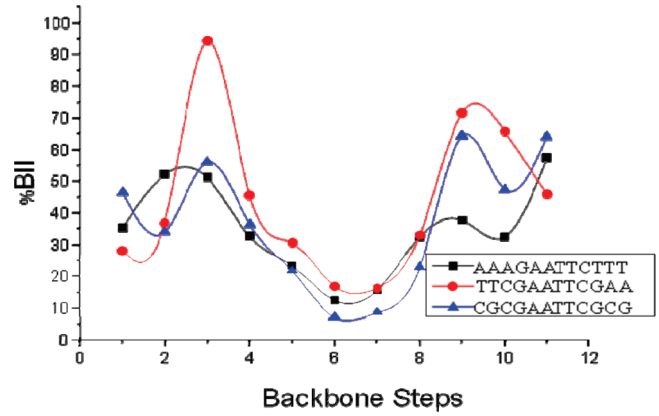

FIGURE 6: Variation in backbone conformation (reported as % BII) as a function of the nucleotide step for *Eco*RI AAA/TTT (black), *Eco*RI TTC/GAA (red), and *Eco*RI CGC/GCG (blue).

backbone conformations across the entire sequence whereas *Eco*RI TTC/GAA shows steps with dramatic percent BII character. This is in agreement with the FTIR data showing a broad resonance at 970 cm^{-1} for *Eco*RI AAA/TTT and resonances at varying backbone conformations for *Eco*RI TTC/GAA. The *Eco*RI CGC/GCG sequence also shows variation in backbone conformation, again agreeing with the variety and breadth of backbone resonances apparent in its FTIR spectrum.

The marker for adenine-thymine base pairs in B-form helices occurs in all samples at 931–928 cm^{-1} (Figure 4). In addition, the deoxyribose peak at 898−893 cm^{-1} is evident in all samples but broadest in the *Eco*RI CGC/GCG sequence. Broad peaks centered between 839 and 835 cm^{-1}, arising from S-type sugar puckers, are evident in three samples. The derivative spectra (Figure 5) show an almost equal population into two S-type markers at 835 and 829 cm^{-1} as well as peaks at 815 and 804 cm^{-1}, indicative of N-type sugars in *Eco*RI CGC/GCG. Again, these results demonstrating varying furanaose structure and the presence of furnanose dynamics are consistent with previous studies of the furanose structure and dynamics in the *Eco*RI CGC/GCG sequence (*34*−*36*). The *Eco*RI AAA/TTT and *Eco*RI TTC/GAA sequences demonstrate peaks at 846, 836, and 804 cm^{-1} and an O4′ *endo* sugar puckering marker at 825 cm^{-1} (*16*).

DISCUSSION

Our findings demonstrate that one can use FTIR and the library of infrared marker bands compiled by Banyay et al. (*23*) to determine the structural and dynamical changes conferred by different flanking sequences. This approach allowed us to examine unique peaks and subtle changes in the spectra of *Eco*RI binding sites that have been flanked by various trinucleotides, and thereby identify local changes in base pairing, base stacking, backbone conformation, glycosidic bond rotation, and sugar puckering in the studied sequences. We included the analysis of the well-studied Dickerson dodecamer in this investigation so that structural features observed via FTIR could be directly compared with previously reported NMR, X-ray, and molecular dynamics investigations of this sequence.

Variations in the nucleotide sequence produce different vibrations within both the direct and indirect readout regions.

Based on the peak assignments in the 1800−1500 cm^{-1} region, most of the nucleotides in the *Eco*RI sequences are double stranded. However, in all dodecamers studied peaks between 1673–1660 and 1608−1600 cm^{-1}, corresponding to a single-stranded guanine nucleotide, are observed. These peaks are even evident in the *Eco*RI AAA/TTT sample that contains only one guanine nucleotide adjacent to the *Eco*RI restriction endonuclease cutting site (AAAG↑AATTCTTT). This finding supports the presence of a splay in the G_4-C_9 base pair in the *Eco*RI cutting site, which was described by Drew et al. and seen in NMR and molecular dynamics studies (*32, 35, 37*) in the crystal structure of the B-form helix. We also observed a number of changes in base stacking that are attributed to the identities and positions of the bases in the flanking sequences of these dodecamers. Moreover, they can be credited to vibrations from the flanking sequences affecting the nucleotides within the GAATTC binding site. For example, in *Eco*RI TTC/GAA, the vibrations of base-paired cytosines occur at 1528 cm^{-1} and shift downfield to 1535 cm^{-1} in *Eco*RI AAA/TTT, a change that may be attributed to a decrease in the percentage of cytosines in the dodecamer. This region does show significant variations in the base stacking patterns of samples with different flanking sequences. This variation in base stacking as a function of flanking sequence is in agreement with the FTIR results of Lindqvist and Graslund (*22*). It has been previously demonstrated that changes in base stacking effects are directly related to backbone conformations (*38*−*40*). The current FTIR results of the backbone resonance breadth and ^{31}P NMR chemical shift data are thus in keeping with the variety of base stacking resonances we observe.

Changes in the endocyclic torsion angles of the furanose ring are evident by studying the 1000−800 cm^{-1} region of the FTIR spectra. Our results show that all sequences contain both N-type and S-type sugars and that the S-type sugars are well resolved into C2′ *endo* and O4′ *endo* peaks. The *Eco*RI CGC/GCG sequence has broad sugar markers that demonstrate a variety of conformations in the derivative spectrum. Such furanose dynamics have been directly observed in the *Eco*RI CGC/GCG sample using solid-state ^2H NMR (*34*). All sequences have the B-form backbone marker at 970 cm^{-1} with the marker in the *Eco*RI TTC/GAA and *Eco*RI CGC/GCG sequences being broad and possessing a backbone resonance at 995 cm^{-1}. The derivative spectra shown in Figure 5 demonstrate various backbone states in these samples. We attribute this to the presence of the BI and BII conformational substates of the backbone (*41, 42*). The breadth of the peak in the *Eco*RI CGC/GCG sequence is in agreement with recent ^{31}P dynamic NMR results in the *Eco*RI CGC/GCG sequence that demonstrate facile backbone dynamics at the C3G4 and C9G10 backbone steps (*31*).

The region between 1250 and 1000 cm^{-1} is sensitive to vibrations along the sugar−phosphate backbone, and these marker bands provide information about the overall conformation of the double helix (*23*). With the vibrations we observed in this region, we have been able to conclude that all *Eco*RI dodecamers exist predominantly in the B-form conformation as they exhibit peaks between 1224 and 1221 cm^{-1}. Notably, these peaks are broad, again suggesting the sampling of different backbone conformations. Despite the fact that the sequences all favor the B-form as their primary backbone conformation, the *Eco*RI AAA/TTT and TTC/GAA sequences exhibit B-, A-, and Z-form conformation markers in the 1500−1250 cm^{-1} region as well as N-type sugar puckering, in the case of the *Eco*RI AAA/TTT sequence. A mixture of these markers can be attributed to fraying of the helix near the ends of the sequences, since the two other regions of interest of the spectra (1250−1000 and 1000−800 cm^{-1}) also exhibit a dominant B-form. However, it is more likely that the flanking sequences of these dodecamers confer rigidity or flexibility to these sequences. For example, the Z-type marker at 1434−1440 cm^{-1} is present in *Eco*RI TTC/GAA. The N-type sugar marker at 1406 cm^{-1} in *Eco*RI AAA/TTT that became much weaker in *Eco*RI TTC/GAA and completely disappeared in *Eco*RI CGC/GCG may indicate that *Eco*RI AAA/TTT and *Eco*RI TTC/GAA tend to exhibit more of the rigid A-form and Z-form conformations whereas *Eco*RI CGC/GCG primarily favors the flexible B-form. The marker for purines in the *anti* conformation appears at 1373 and 1365 cm^{-1}. The strong presence of this peak in *Eco*RI AAA/TTT is likely due to the A-tract flanking sequence. The *Eco*RI TTC/GAA sequence has a broad range of peaks in this region encompassing the *syn* marker at 1323 cm^{-1}. The *Eco*RI CGC/GCG sequence shows far less resolution in this region, again suggesting a more dynamic sugar ring.

Our findings suggest that the *Eco*RI AAA/TTT and *Eco*RI TTC/GAA sequences are more similar to each other than to the *Eco*RI CGC/GCG dodecamer in the 1800−1500 cm^{-1} and 1500−1250 cm^{-1} base-reporting regions. The *Eco*RI CGC/GCG spectrum is much smoother with broader peaks and more poorly defined marker bands than the other sequences. This evidence, along with the observation that local nucleotides in the rigid A-form and Z-forms occur more frequently within the *Eco*RI AAA/TTT and *Eco*RI TTC/GAA sequences, suggests that *Eco*RI CGC/GCG is the most flexible sequence.

In fact, the FTIR regions reporting on the backbone and sugar moieties all show increased breadth for the *Eco*RI CGC/GCG sequence. This is in keeping with the previous ^{31}P and ^2H NMR (*34*) work on *Eco*RI CGC/GCG and the work of Windolph et al., who demonstrated increased cleavage rates of the K130E mutant of *Eco*RI on sequences with "favorable" flanking sequences (*12*). The cocrystal structure of *Eco*RI bound the *Eco*RI CGC/GCG sequence shows that there are no direct base contacts to the flanking sequences (*43*) so these sequences must instead modulate the DNA structure or flexibility. The Windolph study used known algorithms (*44, 45*) to study this modulation and found that there was no real difference in the stacking energies of their starting pool and their "best cut" or "worst cut" pools, indicating that stacking energies are not responsible for the changes in cutting rates. However, they did demonstrate increased flexibility of their best cut pool using the flexibility algorithms of Satchwell and Travers (*46*). We have shown this flexibility in the *Eco*RI CGC/GCG sequence using ^{31}P and ^2H NMR (*34*). Here, we directly demonstrate this increased flexibility as a function of flanking sequence using FTIR.

We have been able to conclude this using FTIR as this methodology provides a "snapshot" of the nucleic acid sequences where sharp resonances are the result of defined conformations while broad ones suggest dynamical averaging on the FTIR time scale. That flanking sequences are able to

affect a sequence's flexibility is in keeping with those of Lefebvre et al. (*18*), who used NMR to study the effect of modified flanking sequences on a CpG step and found that overall the $d(CATCGATG)_2$ sequence was more flexible than the $d(CTTCGAAG)_2$ octamer. Similarly, Cordier et al. (*11*), who used ^1H NMR and ^{31}P NMR, found that A and T tracts of three base pairs on either side of the CRE sequence, $d(GAAAACGTTTTC)_2$, induced rigidity in this sequence as compared to flexibility observed in $d(ATGACGTCAT)_2$. This confirmation of differing flexibility due to flanking sequence has important biological ramifications. It is known from the cocrystal structure of the *Eco*RI restriction endonuclease in complex with its recognition sequence that this protein−DNA interaction utilizes an induced fit mechanism (*47*). The neokinks, distortions in the DNA structure upon complexation, incur an energetic cost, and this cost is presumed to vary based on surrounding sequences (*9*). Thus, the changes in cleavage rates observed by Windolph were suggested to be caused by altered DNA flexibility caused by flanking sequence effects. Here, we have provided direct experimental evidence supporting that hypothesis.

In conclusion, we have shown variations in FTIR spectra that efficiently provide preliminary information about the specific local structural response of the *Eco*RI binding sites to changes in the trinucleotides flanking these sites. In particular, we found that different flanking sequences can (1) promote A- or Z-type structures in primarily B-form DNA sequences, (2) alter backbone and sugar conformations, and (3) promote dynamical averaging of backbone and furanose conformations. While these results are consistent with previous research (*11, 18*), we recognize that this is an underdetermined problem, and we cannot confidently identify the source of IR band shifts. We plan to continue this investigation using our recent ^{31}P dynamic NMR method (*31*) to provide the quantitative dynamic data FTIR cannot yield. However, the application of this method to the well-studied *Eco*RI CGC/GCG sequence provides confirmation for the usefulness of FTIR in demonstrating the flexibility of the DNA backbone. In addition, by using the *Eco*RI CGC/GCG sequence as a control, we have demonstrated that our FTIR results correspond to the structural features determined using other techniques (*31, 32, 34, 35, 48*). Thus, we have shown that FTIR provides a rapid and efficient method for examining a large number of DNA sequences in order to uncover trends and to determine which samples are worthy of further investigation by more quantitative techniques.

The obvious changes in FTIR spectra caused by altering flanking sequences confirm that flanking sequences have an important role in local structure and flexibility in DNA sequences. This is of tremendous importance because it suggests that the local environment of the DNA binding site determines its binding affinity. Understanding this phenomenon underscores the important role of indirect readout in protein−DNA recognition events (*15, 16, 42*).

ACKNOWLEDGMENT

The authors gratefully acknowledge programming assistance from Aaron Skeers, NMR support from Professor Len Mueller and Yi Tian at UC Riverside, and insightful discussions and careful editing by Diana Buckett, Stephanie McCarty, and Sarah Primrose.

REFERENCES

1. Grohima, M. (2005) Influence of DNA stiffness in protein-DNA recognition. *J. Biotechnol. 117*, 137–145.
2. Harrington, R. (1992) DNA curving and bending in protein-DNA recognition. *Mol. Microbiol. 6*, 2549–2555.
3. Narayana, N., Ginell, S. L., Russu, I. M., and Berman, H. M. (1991) Crystal and molecular structure of a DNA fragment: d(CGTGAATTCACG). *Biochemistry 30*, 4449.
4. Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) Crystal structure of trp repressor/operator complex at atomic resolution. *Nature 335*, 321–329.
5. Shakked, Z., Guzikevich-Guerstein, G., F., Frolow, F., D., Rabinovich, D., Joachimiak, A., and Sigler, P. B. (1994) Determinants of repressor/operator recognition from the structure of the trp operator binding site. *Nature 368*, 469–473.
6. Carey, M., and Smale, S. T. (2000) *Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
7. Hogan, M. E., and Austin, R. H. (1987) Importance of DNA stiffness in protein-DNA binding specificity. *Nature 329*, 263–266.
8. Sarai, A., Mazur, J., Nussinov, R., and Jernighan, R. L. (1989) Sequence dependence of DNA conformation flexibility. *Biochemistry 28*, 7842.
9. Lesser, D. R., Kurpiewski, M. R., Waters, T., Connolly, B. A., and Jen-Jacobson, L. (1993) Facilitated distortion of the DNA site enhances EcoRI endonuclease-DNA recognition. *Biochemistry 90*, 7548–7552.
10. Nathan, D., and Crothers, D. M. (2002) Bending and flexibility of methylated and unmethylated *Eco*RI DNA. *J. Mol. Biol. 316*, 7–17.
11. Cordier, C., Marcourt, L., Petitjean, M., and Dodin, G. (1999) Conformation variation of the central CG site in $d(ATGACGTCAT)_2$ and $d(GAAAACGTTTTC)_2$. An NMR, molecular modeling and 3D-homology investigation. *Eur. J. Biochem. 261*, 722–733.
12. Windolph, S., Fritz, A., Oelgeschlager, T., Wolfes, H., and Alves, J. (1997) Sequence context influencing cleavage activity of the K130E mutant of restriction endonuclease *Eco*RI identified by a site selection assay. *Biochemistry 36*, 9478–9485.
13. Riccelli, P. V., Vallone, P. M., Kashin, I., Faldasz, B. D., Lane, M. J., and Benight, A. S. (1999) Thermodynamic, spectroscopic, and equilibrium binding studies of DNA sequence context effects in six 22-base pair deoxyoligonucleotides. *Biochemistry 38*, 11197–11208.
14. Qu, X., Ren, J., Riccellie, P. V., Benight, A. S., and Chaires, J. B. (2003) Enthalpy/entropy compensation: Influence of DNA flanking sequences on the binding of 7-amino actinomycin D to its primary binding site in short DNA duplexes. *Biochemistry 42*, 11960–11967.
15. Wellenzohn, B., Flader, W., Winger, R. H., Hallbrucker, A., Mayer, E., and Liedl, K. R. (2001) Complex of B-DNA with polyamides freezes DNA backbone flexibility. *J. Am. Chem. Soc. 123*, 5044–5049.
16. Banyay, M., and Graslund, A. (2002) Structural effects of cytosine methylation on DNA sugar pucker studied by FTIR. *J. Mol. Biol. 324*, 667–676.
17. Lankas, F., Sponer, J., Langowski, J., and Cheatham, T. E. (2003) DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J. 85*, 2872–2883.
18. Lefebvre, A. O. M., el Antri, S., Monnot, M., Lescot, E., and Fermandjian, S. (1995) Sequence dependent effects of CpG cytosine methylation. A joint ^1H-NMR and ^{31}P-NMR study. *Eur. J. Biochem. 229*, 445–454.
19. Vallone, P. M., and Benight, A. S. (2000) Thermodynamic, spectroscopic, and equilibrium binding studies of DNA sequence context effects in four 40 base pair deoxynucleotides. *Biochemistry 39*, 7835–7846.
20. Siebert, E., Ross, J. B., and Osman, R. (2003) Contribution of opening and bending dynamics to specific recognition of DNA damage. *J. Mol. Biol. 330*, 687–703.
21. Grohima, M. M. (2000) Structure based sequence dependent stiffness scale for trinucleotides: A direct method. *J. Biol. Phys. 26*, 43–50.
22. Lindqvist, M., and Graslund, A. (2001) An FTIR and CD study of the structural effects of G-tract length and sequence context on DNA conformation in solution. *J. Mol. Biol. 314*, 423–432.
23. Banyay, M., Sakar, M., and Graslund, A. (2003) A library of IR bands of nucleic acids in solution. *Biophys. Chem. 104*, 477–488.

24. Pichler, A., Rudisser, S., Winger, R. H., Liedl, K. R., Hallbrucker, A., and Mayer, E. (2000) The role of water in B-DNAs BI to BII conformer substates interconversion: A combined study by calorimetry, FT-IR spectroscopy and computer simulation. *Chem. Phys. 258*, 391–404.
25. Liquier, J., Akhebat, A., Taillandier, E., Ceolin, F., Huynh Dinh, T., and Igolen, J. (1991) Characterization by FTIR spectroscopy of the oligoribonucleotide duplexes r(A-U)$_6$ and r(A-U)$_8$. *Spectrochim. Acta 47*, 177–186.
26. Tsuobi, M. (1969) Application of infrared spectroscopy to structure studies of nucleic acids, in *Applied Spectroscopy Reviews* (Brame, E. G. J., Ed.) pp 45−90, Dekker, New York.
27. Kay, L. E., Keifer, P., and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc. 114*, 10663–10665.
28. Palmer, A. G., Cavanaugh, J., Wright, P. E., and Rance, M. (1991) Sensitivity improvement in proton-detected 2-dimensional heteronuclear correlation NMR-spectroscopy. *J. Magn. Reson. 93*, 151–170.
29. Schleucher, J., Schwendinger, M., Sattler, M., Schmidt, P., Schedletsky, O., Glaser, S. J., Sorensen, O. W., and Griesinger, C. (1994) A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed-field gradients. *J. Biomol. NMR 4*, 301–306.
30. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
31. Tian, Y., Kayatta, M., Shultis, K., Gonzalez, A., Mueller, L. J., and Hatcher, M. E. (2008) ^{31}P NMR investigation of backbone dynamics in DNA binding sites. *J. Phys. Chem. B* (in press).
32. Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., and Dickerson, R. E. (1981) Structure of a B-DNA dodecamer−Conformation and dynamics. 1. *Proc. Natl. Acad. Sci. U.S.A. 78*, 2179–2183.
33. Lee, C., and Cho, M. (2007) Vibrational dynamics of DNA: IV. Vibrational spectroscopic characteristics of A-, B-, and Z-form DNAs. *J. Chem. Phys. 126*, 145102–145110.
34. Hatcher, M. E., Mattiello, D. L., Meints, G. A., Orban, J., and Drobny, G. P. (1998) A solid-state deuterium NMR study of the localized dynamics at the C9pG10 step in the DNA dodecamer [d(CGCGAATTCGCG)]$_2$. *J. Am. Chem. Soc. 120*, 9850–9862.
35. Nerdal, W., Hare, D. R., and Reid, B. R. (1989) Solution structure of the *Eco*RI DNA sequence: Refinement of NMR-derived distance geometry structures by NOESY spectrum back-calculations. *Biochemistry 28*, 10008–10021.
36. Bax, A., and Lerner, L. (1988) Measurement of ^{1}H-^{1}H coupling constants in DNA fragments by 2D NMR. *J. Magn. Reson. 79*, 429–438.
37. Withka, J. M., Swaminathan, S., Srinivasan, J., Beveridge, D. L., and Bolton, P. H. (1996) Toward a dynamical structure of DNA: Comparison of theoretical and experimental NOE intensities. *Science 255*, 597–599.
38. Hartmann, B., Piazzola, D., and Lavery, R. (1993) BI-BII transitions in B-DNA. *Nucleic Acids Res. 21*, 561–568.
39. Heddi, B., Foloppe, N., Hantz, E., and Hartmann, B. (2007) The DNA structure responds differently to physiological concentrations of K^+ or Na^+. *J. Mol. Biol. 368*, 1403–1411.
40. Djuranovic, D., and Hartmann, B. (2003) Conformational characteristics and correlations in crystal structures of nucleic acid oligonucleotides: Evidence for sub-states. *J. Biomol. Struct. Dyn. 20*, 771–788.
41. Rudisser, S., Hallbrucker, A., and Mayer, E. (1997) B-DNA's conformational substates revealed by Fourier transform infrared difference spectroscopy. *J. Am. Chem. Soc. 119*, 12251–12256.
42. Flader, W., Wellensohn, B., Winger, R. H., Hallbrucker, A., Mayer, E., and Liedl, K. R. (2001) B1→BII substate transitions induce changes in the hydration of B-DNA, potentially mediating signal transduction from the minor to major groove. *J. Phys. Chem. 105*, 10379–10387.
43. Kim, Y., Grable, J. C., Love, R., Greene, P. J., and Rosenberg, J. M. (1990) *Science 249*, 1307–1309.
44. Santa Lucia, J., Allawi, H. T., and Senevirante, P. A. (1996) *Biochemistry 35*, 3555–2562.
45. Delcourt, S. G., and Blake, R. D. (1991) *J. Biol. Chem. 266*, 15160–15169.
46. Satchwell, S. C., and Travers, A. A. (1989) *EMBO J. 8*, 229–238.
47. McClarin, J. A., Frederick, C. A., Wang, B. C., Boyer, H. W., Grable, J. C., and Rosenberg, J. M. (1986) *Science 234*, 1526–1541.
48. Shui, X. Q., McFail-Isom, L., Hu, G. G., and Williams, L. D. (1998) The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry 37*, 8341–8355.

BI8015235