Environmentally Induced Conformational Changes in B-Type DNA: Comparison of the Conformation of the Oligonucleotide d(TCGCGAATTCGCG) in Solution and in Its Crystalline Complex with the Restriction Nuclease *Eco*RI†

Gerald A. Thomas, William L. Kubasek, and Warner L. Peticolas*

Department of Chemistry and Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403

Patricia Greene

Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California 94143

John Grable and John M. Rosenberg

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Received April 29, 1988; Revised Manuscript Received August 15, 1988

ABSTRACT: Raman spectroscopic analysis of the secondary structure of the crystalline restriction endonuclease *Eco*RI, the oligonucleotide d(TCGCGAATTCGCG) in solution, and the corresponding crystalline *Eco*RI–oligonucleotide complex reveals structural differences between the complexed and uncomplexed protein and oligonucleotide components that appear to be linked to complex formation. Structural differences that are spectroscopically identified include (1) an increase in the population of furanose rings adopting the C3′-endo conformation and (2) spectroscopically observed changes in base stacking which are probably associated with the crystallographically observed distortion of the phosphate backbone about positions C(3)-G(4) and C(9)-G(10) and unwinding between the symmetry-related segments GAA-TTC which make up the central recognition core (McClarin et al., 1986). Changes in base stacking due to distortions and unwinding along the oligonucleotide result in differences in the base vibrational region between the spectra of the complex and the oligonucleotide in solution. The spectroscopic analysis indicates that the C2′-endo population is similar for the oligonucleotide in solution and in the complex. The additional C3′-endo population in the complex appears to arise from the conversion of rings adopting alternative conformations such as C1′-exo and O1′-endo. Analysis of the vibrational bands derived from guanine indicates that the population of guanine residues associated with furanose rings in a C2′-endo conformation is similar for the oligonucleotide in solution and in the crystalline complex. This implies that the increase in C3′-endo population is not associated with guanine residues. Large conformational distortions such as those observed in the crystal are not observed for the oligonucleotide in solution. Furthermore, Raman evidence is presented that these distortions are not observed in either the crystal or the solution of the oligomer d(CGCGAATTCGCG). These data suggest that static distortions such as those observed in the crystal are not employed for initial sequence recognition. They appear either as a secondary response to interaction with the protein or as transient fluctuations which exist at a very low level in solution.

The phenomenal efficiency by which proteins recognize and interact with a specific nucleic acid sequence from among the multitude of potential binding sequences present in cellular DNA not only is remarkable but also is crucial to normal cellular development and activity. Such recognition and binding involve an intricate set of molecular interactions and equilibria which are governed by the chemical identity and organization of the binding species and by the conformational and hydrogen-bonding characteristics of the recognition sequence. A comprehensive understanding of the mechanism underlying sequence-specific recognition requires a detailed knowledge of the molecular and structural interactions which are present at the protein–DNA interface. Recently, the crystal structures of several proteins which exhibit sequence-specific recognition have been solved including the Cro and CI repressor proteins from Coliphage λ (Anderson et al., 1981, 1982; Pabo & Lewis, 1982), the Ecoli catabolite gene activator protein (CAP) (Mckay & Steitz, 1981), the tryptophan repressor (Schevitz et al., 1985), Coliphage 434 repressor (Anderson et al., 1985), and the endonuclease *Eco*RI (Frederick et al., 1984; McClarin et al., 1986). The first four protein structures were solved in the absence of DNA and a common α-helix–turn–α-helix structural motif supporting the DNA binding site has been identified. Preliminary analysis of co-crystals of Cro with the oligonucleotide d(TATCACCGCGGGTAGATA) suggests that the α-helix–turn–α-helix motif is preserved upon DNA binding (Brennan and Matthews, personal communication). General features of the DNA–434 repressor complex observed at 7-Å resolution also support a similar mode of DNA binding (Anderson et al., 1985). The structural similarities in the mode of DNA binding observed for the CAP activator protein and the Cro, CI, 434, and tryptophan repressor proteins suggest that these proteins may belong to a common class of DNA recognition proteins.

On the other hand, a different structural variant is employed by the endonuclease *Eco*RI to achieve DNA binding. The endonuclease is a small protein (31 065 daltons, 276 amino acids) which exhibits highly specific recognition of the double-stranded sequence d(GAATTC) (Greene et al., 1981;

† This work was supported by grants from the National Institutes of Health (GM15547, GM33825, and GM25671).

0006-2960/89/0428-2001$01.50/0

Newmann et al., 1981). The protein forms highly stable catalytically active dimers in solution, and binding to the recognition sequence in the presence of Mg^{2+} results in cleavage of the phosphodiester linkage between the guanylic and adenylic acid residues (Modrich & Zabel, 1976; Jen-Jacobson et al., 1983). The process cumulates with release of a nucleic acid with a single-stranded end of the sequence 5′-p-AATT. In the absence of Mg^{2+}, protein binding remains specific with a dissociation constant estimated at 10^{-11} though stand cleavage does not occur (Modrich, 1979; Halford & Johnson, 1980; Rosenberg et al., 1981; Jack et al., 1981). The high degree of specificity in binding in the absence of Mg^{2+} has been eploited in the preparation of stable crystals of the endonuclease and the oligonucleotide d-(TCGCGAATTCGCG). The crystalline complex consists of two enzyme subunits and one double helix (Frederick et al., 1984; McClarin et al., 1986). The enzyme subunits are arranged symmetrically about the recognition site (GAATTC) located at the center of the self-complementary oligonucleotide. At the reported resolution of 3 Å, the secondary structural elements of the protein are clear, and two-thirds of the amino acid side chains can be unambiguously identified. Each protein is organized into a five-stranded β-sheet surrounded on each side by α-helices. Direct interaction between the bases of the recognition sequence and the protein involves the amino acid side chains within α-helical modules. Even so, the superstructure or structural motif supporting DNA recognition and strand scission appears to be provided by the β-sheet organization in the protein.

One of the more remarkable features of the complex is the presence within the oligonucleotide of three abrupt departures from helical (screw) symmetry termed neokinks (Frederick et al., 1984; McClarin et al., 1986) that result in segmentation of the oligonucleotide into symmetry-related trimer units, d(CGC-GAA-TTC-GCG). The *Eco*RI recognition sequence spans the two internal segments. These internal segments are separated by an abrupt unwinding termed the type I neokink. The two symmetrically related terminal segments are linked to the internal segments by distortions designated type II neokinks.

One central concept of the analysis of the neokinks is the assumption that they are induced by the formation of the DNA–protein complex. Indeed, the prefix "neo" is reserved for that situation (McClarin et al., 1986)—an intrinsically stable distortion would simply be refered to as a kink. Therefore, a careful comparison between complexed and uncomplexed states of the DNA is important. The assumption that the neokinks were induced by complex formation was based on a comparison with the crystal structure of the related B-form dodecamer d(CGCGAATTCGCG) that was determined by Dickerson and co-workers (Drew et al., 1981; Dickerson & Drew, 1981). Here, the global structure of the dodecamer is that of a smoothly bent B-form double helix. Locally, the double helix is characterized by severe variation in furanose conformation, helical roll, pitch, base pairs per turn, and base stacking.

Recently, we have reported that Raman spectroscopic analysis indicates that these local variations appear to relax in solution, suggesting that solvation and/or crystal packing plays a role in the modulation of the local distortions of the double helix of d(CGCGAATTCGCG) (Thomas et al., 1987). The general differences encountered in going from the crystal to the solution conformation outlined in our previous report include a reduction in the A–like character of the C,G resiudes which flank the central AATT sequence that is the *Eco*RI recognition site, redistribution of furanose conformations resulting in changes in the C3′-endo and C2′-endo populations, and changes in the base stacking organization about the thymine residues. Subsequent to our initial report of a difference in the crystal and solution Raman spectra of the dodecamer d(CGCGAATTCGCG), a paper has appeared using a different dodecamer, d(CGCAAATTTGCG) (Benevides et al., 1988). These authors also report differences in the secondary structure of the crystalline oligonucleotide and the oligonucleotide in solution. The appearance of two reports of conformational variation in the B forms of different DNAs in going from the crystal to solution state suggests that this effect may be a general one. The question remains, however, as to what effect the interaction of a protein such as *Eco*RI will have on DNA conformation. It therefore becomes imperative to compare the conformation of the DNA in the *Eco*RI complex with that in solution as well.

With the aim of providing additional structural details concerning the *Eco*RI–oligomer d(TCGCGAATTCGCG) complex, a comparative Raman spectroscopic investigation of the crystalline complex and the oligonucleotide in solution was initiated. The spectroscopic technique itself has been applied to the analysis of the secondary structure of oligomers and polymers of defined sequence and DNA–protein complexes including DNA containing viruses and sperm cells (Peticolas et al., 1987; Thomas, 1987; Thomas et al., 1986; Kubasek et al., 1986). The technique provides vibrational information or data which can be used to determine and probe the global as well as local secondary structure of nucleic acids. Specific information available from a static Raman spectroscopic analysis includes identification of global A, B, and Z forms, determination of relative base stacking organization, and quantitative determination of the distribution of furanose conformations which can under favorable circumstances lead to the identification of the furanose conformation of individual residues. A particular advantage of the technique is that it can be applied equally to an investigation of the solution and crystalline states. This provides a means by which environmental and packing effects can be identified and separated from sequential and binding influences. This advanatage has been recently exploited in an analysis of the conformation of 14 oligonucleotides of defined sequence in the crystalline and solution states (Wang et al., 1987a,b). The present report is concerned with a comparative analysis between three different species; the crystalline protein–oligonucleotide complex, the pure crystalline protein, and the oligonucleotide in solution. The principle objective of the report is to facilitate a detailed comparative study of these three forms.

Materials and Methods

Crystals of the *Eco*RI–d(TCGCGAATTCGCG) complex and the pure protein were prepared as described previously (Frederick et al., 1984; McClarin et al., 1986). Crystals were mounted in quartz X-ray-quality capillary tubes and sealed. Capillary tubes also contained a small portion of crystallization mother liquor to prevent desiccation of the crystals. Raman spectra of the crystals were acquired by using an intracavity laser Raman microscope. The apparatus has been described in detail elsewhere (Wang et al., 1987b; Patapoff et al., 1988). The basic components of the system include a Spectra Physics 171 argon ion laser, a Zeiss UEM Universal microscope, a Spex triplemate, and a Tracor Northern optical multichannel analyzer. An excitation wavelength of 514.5 nm was employed for all samples. On average, 20 mW of laser power measured at the source was used. No apparent damage to the crystalline samples was noted over the duration of data collection. Sa-

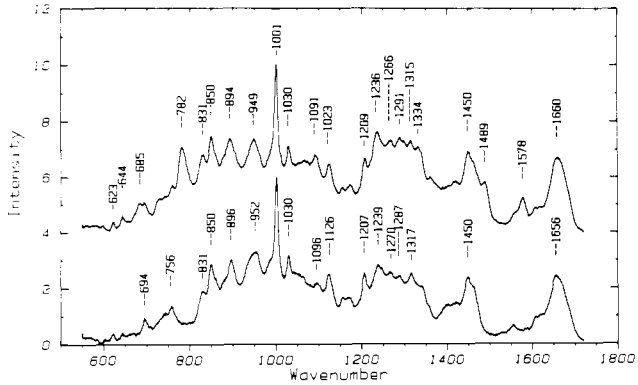

FIGURE 1: Raman spectrum of crystals containing the oligonucleotide d(TCGCGAATTCGCG) bound to the restriction endonuclease *Eco*RI (top). Raman spectrum of crystals of the restriction endonuclease *Eco*RI in the absence of oligonucleotide (bottom).

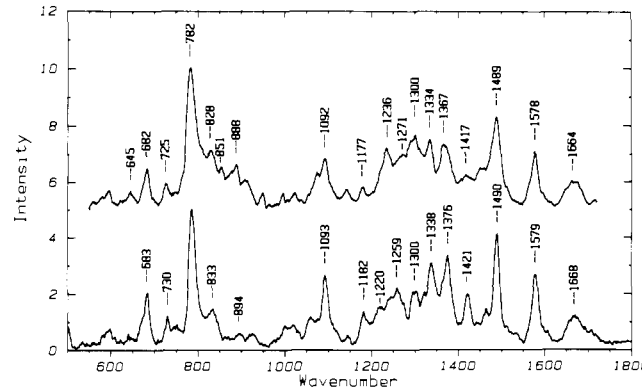

FIGURE 2: (Top) Raman spectrum of the oligonucleotide d-(TCGCGAATTCGCG) bound to the restricion endonuclease *Eco*RI. Protein components have been removed by computer subtraction of the spectrum of the endonuclease crystals (Figure 1, bottom) from the spectrum of the crystalline protein–oligonucleotide complex (Figure 1, top). Normalization of the two spectra in Figure 1 for subtraction was performed by reference to the strong band at 1001 cm^{-1} assigned to phenylalanine. (Bottom) Raman spectrum of the oligonucleotide d(TCGCGAATTCGCG) in a 0.5 M NaCl solution, pH 7, 2 °C.

tisfactory signal to noise levels were obtained after 500 scans or 5 min total time. Solution samples were prepared by dissolving the oligonucleotide in a 0.5 M NaCl solution, pH 7, to a concentration of 2% or 20 mg/mL. Solution spectra were recorded at 2 °C.

Spectral regions warranting detailed analysis were fit to a minimum sum of Lorentzians by a nonlinear least-squares procedure. The analysis provides a means for the isolation and identification of bands which are difficult to visualize due to spectral congestion. The details of the procedure have been described previously (Patapoff et al., 1988). The general process includes generation of a calculated spectrum from a sum of Lorentzian bands followed by minimization of the differences between the calculated and original data by successive iteration and adjustment of position, area, and half-width parameters. The process is terminated when agreement between the calculated and initial data is reached. Adequacy of agreement was determined by visual inspection and by minimization of the standard deviation between the calculated and the initial data. Overall visual differences between the calculated and initial data were required to less than 1%. The number of Lorentzian bands included in the fitting procedure was restricted to a minimum. The minimum number of bands restriction-enhances the uniqueness of each fitting solution while minimizing the possibility of multiple solutions. The method of analysis has the advantage that the raw data are preserved and not subject to filtering and truncation.

Results and Discussion

The Raman spectrum of crystals of the *Eco*RI–oligonucleotide complex and of the pure protein have been acquired utilizing a intracavity laser microscope apparatus capable of examining material a few microns in size. The Raman spectrum of the complex and the pure protein are displayed in the top and bottom spectra of Figure 1, respectively. The complex spectrum is dominated by contributions from the protein, though bands attributed to the oligonucleotide are readily identified. To facilitate analysis of the structure of the oligonucleotide in the complex, the Raman spectrum of the oligonucleotide can be obtained by computer subtraction of the pure protein spectrum from the protein–oligonucleotide complex spectrum. The resulting difference spectrum is displayed in Figure 2 (top). Subtraction was performed by normalization to the strong band at 1001 cm^{-1} assigned to phenylalanine. The difference spectrum contains almost exclusively bands attributed to the oligonucleotide. Even so, the spectrum may contain potential contributions arising from the protein which are due to structural differences which may exist between the protein in the complex and the pure protein. In this regard, one must proceed with some caution in the analysis of the difference spectrum, and care must be taken in the identification of oligonucleotide and protein components.

Global or Average Oligonucleotide Secondary Structure. Changes in the oligonucleotide structure due to accommodation of the endonuclease can be ascertained by a comparative analysis of the spectrum of the complexed oligonucleotide to the spectrum of the pure oligomer in solution (Figure 2, bottom). The pure oligonucleotide in solution with the exception of the possible existence of fraying or unstacking of the unpaired ending thymine residues is expected to adopt an average B-form structure which is similar to that observed for the oligonucleotide d(CGCGAATTCGCG) in solution. The general spectral features of the oligonucleotide d-(TCGCGAATTCGCG) in the complex (Figure 2, top) and in solution (Figure 2, bottom) correspond to those of an average B-form structure. The two most easily identified spectral indicators of a B form are the presence of a phosphate vibration at 1093 cm^{-1} and a furanose phosphate vibration between 828 and 840 cm^{-1} (Erfurth et al., 1972, 1975; Goodman & Brahms, 1978; Benevides & Thomas, 1983; Thomas & Peticolas, 1983). Both bands are observed in the spectra of the complexed and uncomplexed oligomer. The 830 cm^{-1} band is assigned to a furanose phosphate backbone vibration of the C2′-endo furanose conformation. The C2′-endo furanose conformation is associated with the classical homogeneous structure of B DNA. The 828–840 cm^{-1} band is usually broad and of moderate intensity in the spectra of natural DNA in solution where the B form is dominant. In contrast, a strong sharp backbone vibration between 800 and 815 cm^{-1} is observed for the C3′-endo furanose population found in A-form nucleic acids. The PO^{2-} vibration is observed at 1100 cm^{-1} for A-form DNA and shifts to 1095 cm^{-1} for B-form DNA. The presence of bands at 828 and 1093 cm^{-1} in conjunction with the overall spectral profile of the oligomer suggests that on the average a B-type helix is preserved for both the solution and complexed forms of the oligomer. Though the Raman spectra of the complexed and uncomplexed oligomer at a first glance appear similar, close inspection reveals very distinct and significant spectral differences.

Local Variations in Oligonucleotide Secondary Structure. (A) C3′-Endo Furanose Population. Comparison of the spectra in Figure 2 indicates significant structural perturbations

Table I: Comparison of the Raman Intensity of Selected Bands Observed for the Oligomer d(TCGCGAATTCGCG) and the Crystalline *Eco*RI/d(TCGCGAATTCGCG) Complex[a]

complex (cm^{-1})	change	solution (cm^{-1})	assignment
645	dec	644	A, G, BK
672	inc	669	T, G
682		683	G
726		730	A
745	dec, shift 6 cm^{-1}	751	T
780	dec	783	C
788	inc	791	T, BK(O-P-O)
807	dec	812	C3′-endo BK
1236	dec	sh	T
nr	inc	1259	C
1299	dec	1299	C, A
1334	shift 4 cm^{-1}	1338	G, A
1367		nr	G
nr	inc	1376	T
1417	inc	1421	G, A, BK

[a] Differences in band intensity are described as an intensity change proceeding from the crystalline complex state to the solution state. Abbreviations: nr, not resovable; inc, intensity increase; dec, intensity decrease; sh, shoulder; A, adenine; G, guanine; C, cytosine; T, thymine; BK, deoxyribose phosphate backbone.

and accommodations occur upon formation of the protein–nucleic acid complex. The structural perturbations which occur give rise to the spectroscopic differences listed in Table I and displayed graphically in Figure 3. These differences between the bound and unbound oligomer appear to be localized within the regions from 600 to 1000 cm^{-1} and from 1200 to 1500 cm^{-1}. The 1200–1500 cm^{-1} region is composed principally of base vibrations many of which exhibit appreciable sensitivity to base stacking, hydrogen bonding, and positional organization within the helix. The 600–1000 cm^{-1} region is composed of base vibrations and conformationally sensitive vibrations of the furanose phosphate backbone. The 600–1000 cm^{-1} region of the spectrum of the complexed oligomer is visually quite different than the corresponding spectrum of the oligomer in solution particularly with regard to the presence of additional bands at 851 and 888 cm^{-1}. Spectrally, the 600–900 cm^{-1} region is quite congested, and this region can be examined in greater detail if the individual bands which contribute to the intensity of this region are resolved by band shape analysis. The region from 625 to 975 cm^{-1} for both the complexed and uncomplexed oligonucleotides has been fit to the least possible number of Lorentzians by the use of a nonlinear least-squares program. Calculated curves, initial data, and individual Lorentzian contributions are contained in the top and bottom panels of Figure 4 for the complexed and uncomplexed oligomer, respectively. A weak band at 812 cm^{-1} is observed in the curved resolved spectrum of the oligomer in solution. A band of similar profile and intensity is also observed for the dodecamer d(CGCGAATTCGCG) in solution. The presence of a band between 800 and 815 cm^{-1} has been generally attributed to furanose rings in a C3′-endo conformation (Erfurth et al., 1972, 1975; Goodman & Brahms, 1978; Benevides & Thomas, 1983; Thomas & Peticolas, 1983). The C3′-endo ring conformation is associated with A-form nucleic acids. A band in this region has recently been observed for a number of DNAs of heterogeneous sequence (Wartell & Harrel, 1986). The band at 812 cm^{-1} in solution shifts to 807 cm^{-1} in the complex. The intensity of the band is also larger in the complex. The intensity increase indicates an increase in the population of furanose rings in the C3′-endo conformation. A extremely small band at 807 cm^{-1} is resolved by curve analysis in the spectrum of the related crystalline dodecamer d(CGCGAATTCGCG) (Thomas et al., 1987).

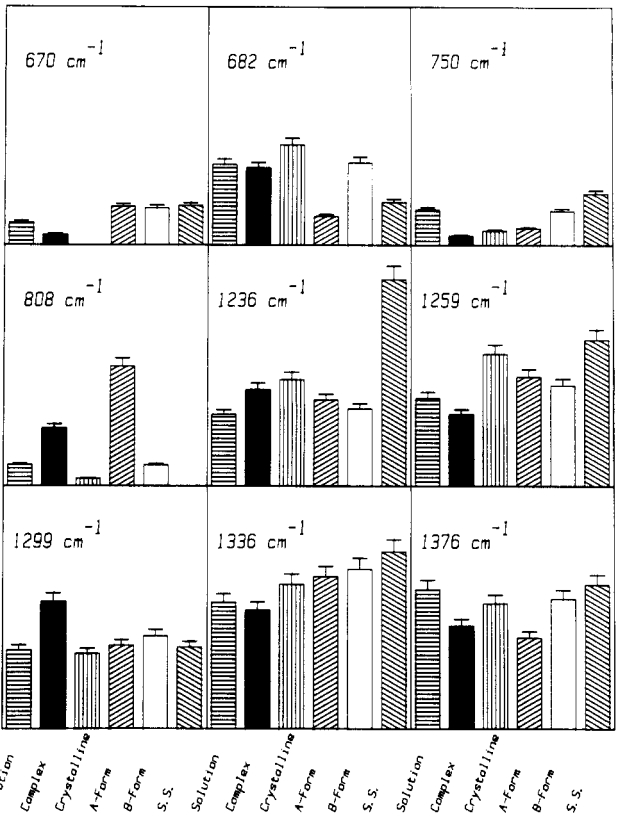

FIGURE 3: Graphical comparison of the relative intensity (increasing vertically) of selected conformationally sensitive Raman bands for the following species: d(TCGCGAATTCGCG) in solution, d(TCGCGAATTCGCG)–*Eco*RI complex, and crystalline d(CGCGAATTCGCG). Maximum error limits are represented by horizontal lines at the top of the bars. The intensity profile of selected Raman bands for the A, B, and single-stranded forms was constructed by examination of the Raman spectra of the corresponding forms for poly[d(A-T)], poly[d(G)]·poly[d(C)], poly[d(G-C)], and calf thymus DNA. The behavior of the 1299 cm^{-1} band is base composition dependent. In polymers containing high adenine content, the band appears at 1304 cm^{-1} and increases in intensity in going from the B to the A form. In polymers of high cytosine content, the band appears at 1298 cm^{-1} and decreases in intensity upon going from the B to the A form. The oligomers examined have a high G-C content (66%), and this is reflected in the intensity profile.

The C3′-endo population of the complexed oligomer appears to be larger than that of either the oligomer or the corresponding dodecamer in solution. The complexed and uncomplexed oligomer d(TCGCGAATTCGCG) as well as the dodecamer d(CGCGAATTCGCG) in solution exhibit an appreciably enhanced C3′-endo population over that observed for the crystalline dodecamer. The relative order of the size of the C3′- endo furanose population is then d(TCGCGAATTCGCG)$_{complex}$ > d(TCGCGAATTCGCG)$_{solution}$ = d(CGCGAATTCGCG)$_{solution}$ > d(CGCGAATTCGCTg)$_{crystalline}$ (see Figure 3).

The similarity in the size of the C3′- endo population for the oligomers d(CGCGAATTCGCG) and d(TCGCGAATTCGCG) in solution is not surprising considering that the double-stranded regions of both oligomers are identical. The size of the C3′-endo population observed in solution is comparable to that observed for polymers of heterogeneous sequences but of similar base composition. Potential energy calculations suggest that the energy barrier preventing interconversion between the C2′-endo and C3′-endo conformations for deoxyfuranose rings is fairly small and adoption of a sizeable C3′-endo population in solution is reasonable (Levitt & Warshel, 1978; Olsen & Sussman, 1982). For the crystalline dodecamer d(CGCGAATTCGCG), the

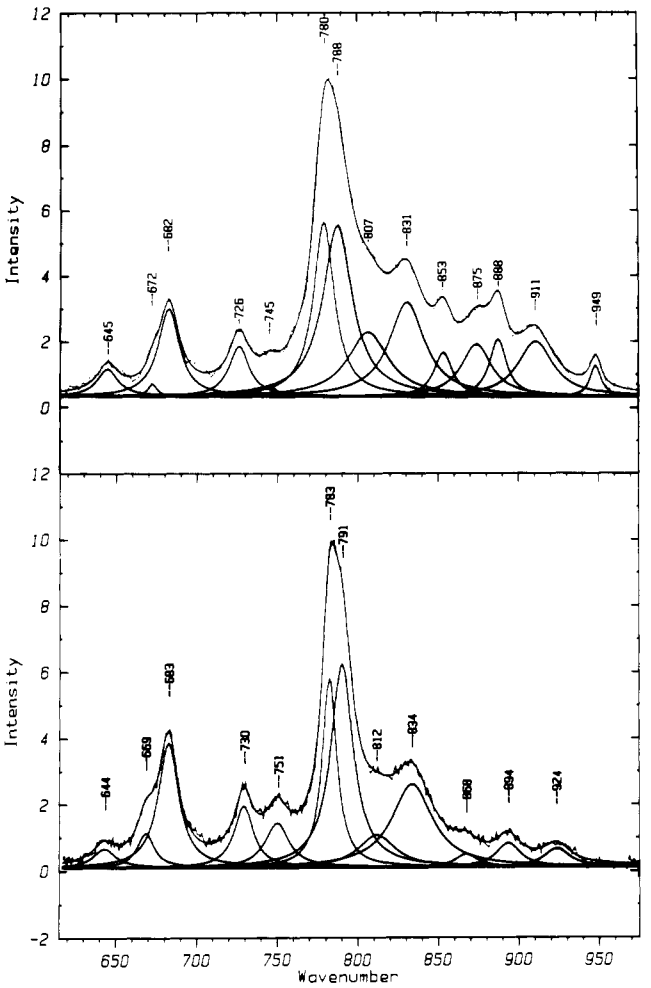

FIGURE 4: (Top) Raman spectrum of the oligonucleotide d-(TCGCGAATTCGCG) bound to the restriction endonuclease *Eco*RI in the spectral region from 615 to 975 cm^{-1}. Protein components have been removed by computer subtraction. The region has been fit to the minimum number of Lorentzian bands, and the results of the fit are displayed as well. The figure contains the calculated curve superimposed upon the original data. Individual band components are displayed. (Bottom) Results of Lorentzian band fit for the region from 615 to 975 cm^{-1} for the oligonucleotide d(TCGCGAATTCGCG) in solution.

number of furanose residues adopting the C3′-endo and O1′-endo conformations is severely limited. In the crystalline state, a single residue in the C3′-endo conformation at the terminal G(12) position and three residues in the O1′- endo conformations have been identified (Drew et al., 1981; Dickerson & Drew, 1981). The O1′-endo furanose conformation occupies a position in conformational space near the top of the potential energy barrier between the energy minimum C2′-endo and C3′-endo conformations. The almost negligible C3′-endo population observed for the crystalline dodecamer appears to be a result of environmental and packing forces. In the absence of external forces, i.e., solvation and intermolecular interactions between neighboring double strands, the C3′-endo population increases (Thomas et al., 1987). The different conformational populations observed for the dodecamer in the crystalline and solution state suggest that the partitioning of furanose rings among the available conformations is determined by an intricate interplay between the primary sequence of the oligomer and environmental constraints.

Other conformational differences between the crystalline and solution dodecamer, d(CGCGAATTCGCG), have also been identified. These differences can be appreciated by

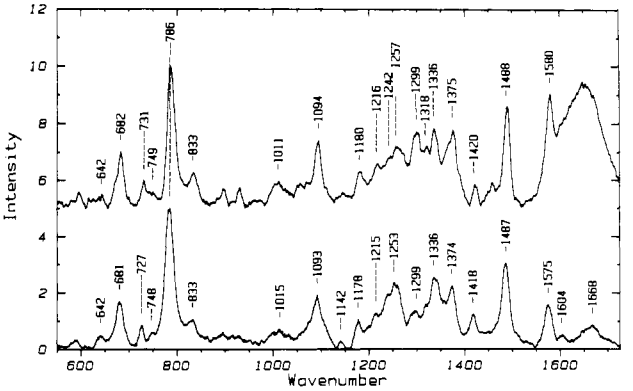

FIGURE 5: Raman spectra of crystals of d(CGCGAATTCGCG) (bottom) and d(CGCGAATTCGCG) in 0.5 M NaCl solution, pH 7.0 (top).

Table II: Comparison of the Raman Intensity of Selected Bands Observed for the Oligomer d(CGCGAATTCG) in the Solution and Crystalline States[a]

complex (cm^{-1})	change	solution (cm^{-1})	assignment
641	dec	646	A, G, BK
nr		672	T, G
681		683	G
727		731	A
748	inc	751	T
780	inc	784	C
788	dec	791	T, BK(O–P–O)
808	inc	810	C3′-endo BK
832	s.inc	836	C2′-endo BK
1093	s.inc	1094	BK(PO^{2-})
1215		1216	T
1242		1242	T
1253	dec	1257	C
1299	inc	1299	C, A
nr		1318	G
1336	s.dec	1336	G, A

[a] Differences in band intensity are described as an intensity change proceeding from the crystalline state to the solution state. Abbreviations: nr, not resovable; s.inc., slight intensity increase; inc, intensity increase; dec, intensity decrease; sh, shoulder; A, adenine; G, guanine; C, cytosine; T, thymine; BK, deoxyribose phosphate backbone.

reference to Figure 5. In this figure, the bottom spectrum was acquired from a single crystal of d(CGCGAATTCGCG) grown in a manner identical with that reported by Wing et al. (1980). The top spectrum is that of the dodecamer in solution. One can see from inspection of the figure distinct differences in the Raman spectrum of the crystalline and solution dodecamer. Many of the bands which differ in the two spectra have been previously shown to be sensitive to DNA conformation. Table II lists the principle bands which change in going from the crystal to solution state. A detailed analysis of these spectra is the subject of another report, and the general findings have been summarized previously (Thomas et al., 1987). The spectra are presented here to provide a useful reference for comparison between the crystalline dodecamer and the crystalline *Eco*RI–13-mer complex.

An estimation of the C3′-endo population in the complex can be crudely made based upon the intensity the 807–814 cm^{-1} band. The intensity of the 807–814 cm^{-1} band relative to the PO^{2-} stretch at 1100 cm^{-1} varies from a value of 1.62 to 2 for A-form nucleic acids of heterogeneous sequence (Brown et al., 1972; Thomas & Hartman, 1973). A value of 2.02 is observed for the crystalline A form of d(GGTATACC) (Wang et al., 1987b; Shaked et al., 1983). If one assumes that the intensity of the 1093 cm^{-1} band does not vary between the complexed and uncomplexed oligomer, then one can arrive at

an upper estimate of the population of furanose rings in a C3′-endo conformation. Taking a value for 2 for a 100% C3′-endo population, then one arrives at an upper estimate of seven to eight residues of the complexed oligonucleotide in the C3′-endo conformation. The estimated C3′-endo population for the oligomer in solution is three to four residues.

The environmental and packing forces placed upon the oligomer d(TCGCGAATCGCG) in the crystalline *Eco*RI complex and the crystalline dodecamer are quite different. Individual protein–DNA complexes are arranged such that the DNA double helices pack end to end. The unpaired thymine residues of neighboring double helices stack on one another, leading to a continuous series of stacked bases running through the crystal (Frederick et al., 1984; McClarin et al., 1986). Thus, the DNA–DNA contacts are limited to the ends of the double helices with extensive DNA–protein contacts lateral to that axis. This contrasts to the crystal packing of the related pure oligonucleotide d(CGCGAATTCGCG) where individual molecules are staggered with each molecule overlapping by three residues with its neighbors above and below (Drew et al., 1981; Dickerson & Drew, 1981). In the latter case, there are significant lateral DNA–DNA contacts and very different interactions in the direction of the helix axis as well. Therefore, protein–DNA contacts dominate the environment of the oligonucleotide within the crystalline protein–DNA complex and appear to be responsible for the elevated C3′-endo population identified spectroscopically in the *Eco*RI–DNA complex.

(B) C2′-Endo Furanose Population and Possible Protein Conformational Changes. As mentioned earlier, both the complexed and uncomplexed oligomers exhibit a band at 828–833 cm^{-1} which indicates a sizable C2′-endo population. The bands are similar in both cases with the intensity of the complexed oligomer slightly larger than for the uncomplexed oligomer. In the case of the complexed oligomer, analysis of the band is complicated by possible protein contributions to this band. In the spectrum of the pure protein in the absence of the oligomer, a very strong tyrosine doublet at 828 and 850 cm^{-1} is observed. The doublet is sensitive to the environment of the tyrosine residues. The relative intensity of the doublet changes in response to alterations in hydrogen bonding, ionization, and dipole interactions (Siamwiza et al., 1975). In the Raman spectrum of proteins and peptides, the tyrosine bands at 828 and 851 cm^{-1} are accompanied by a band at 644 cm^{-1}. Subtraction of the pure protein from the complex (Figures 2 and 4, top spectra) reveals the presence of bands at 641, 628, and 651 cm^{-1}. Bands at 644 and 830 cm^{-1} are also observed for the pure oligomer so that unambiguous separation of protein and oligonucleotide components for these two bands is difficult. If the 850 cm^{-1} band is taken as due solely to tyrosine, then the residual intensity observed in the complexed oligomer spectrum corresponds to one to two tyrosine residues. Incardonia et al. (1987) have proposed that the ratio $I(644)/[I(828) + I(859)]$ is constant at a value of 0.25 for globular proteins. The ratio can then be used to estimate the maximum tyrosine contribution to the 828 cm^{-1} band of the complex if the tyrosine intensity at 644 and 851 cm^{-1} can be determine accurately. If the intensity increase in the 641 cm^{-1} band observed for the complexation of the oligomer is assigned entirely to the presence of residual tyrosines, then the maximum tyrosine contribution to the band at 828 cm^{-1} can be estimated. On the basis of this analysis, the tyrosine contribution to the intensity at 828 cm^{-1} is estimated at 5–10%.

The residual intensity at 851 cm^{-1} can be attributed either to changes in tyrosine environment or to inadequate subtraction of the protein from the complex spectrum. Elimination of the band at 851 cm^{-1} results in creation of a negative band at 1001 cm^{-1} assigned to phenylalanine. The phenylalanine band is expected to exhibit minimal environmental dependence in comparison to the tyrosine band at 851 cm^{-1} which is known to respond dramatically to changes in environment and hydrogen bonding. Consequently, selection of the 1001 cm^{-1} band as a standard for subtraction is more favorable than the band at 851 cm^{-1}, and oversubtraction of the 1001 cm^{-1} band is unwarranted. On this basis, the residual intensity is likely to be due to changes within the protein. It should be pointed out that this cannot be established unequivocally at this time. The calculated 828:850 intensity ratio for the residual tyrosines is 2:10.

The residual tyrosine intensity could arise from changes in the tyrosine environment. The 828:850 intensity ratio indicates an increase in the strength of hydrogen bonding in which the phenolic oxygen of the tyrosine ring acts as a proton acceptor. In principle, this could occur as a result of changes in tyrosine positioning within the protein or by direct interaction with the oligonucleotide although the structure suggests the former. The amino-terminal residues 1–20 and residues 170–192 form an armlike structure that wraps around the DNA (McClarin et al., 1986). Due to the 2-fold symmetry of the complex, there are two arms, each of which interacts with the oligonucleotide, partially encircling the DNA. Selective proteolytic removal of portions of the amino terminus suggests that the nonspecific contacts between the DNA and the terminal residues within the arm are required for catalytic activity but not binding (Jen-Jacobson et al., 1986). The manner in which the DNA is encircled by the two protein arms indicates that entering and exiting of the binding site by the oligomer are accompanied by movement and/or conformational change within the protein. This movement may be responsible for the change in tyrosine intensity between the pure protein and the complex. A possible scenario is that repositioning of the arm to accommodate entering and securing of the oligonucleotide may alter the environment of tyrosine-193 which is one residue removed from the β-hairpin portions of the arm.

The intensity of the 828 cm^{-1} band remaining, after elimination of possible protein contributions in the manner described above, can be attributed to the oligonucleotide within the complex. The area of the band in the complex is $10 \pm 2\%$ larger than observed in solution. The slight increase in area is accompanied by a 12% reduction in bandwidth from a half-width of 15.3 cm^{-1} to 13.2 cm^{-1}. Retention of the 828 cm^{-1} intensity by the oligomer in the crystalline complex indicates that the C2′-endo population remains fairly fixed in the presence and absence of the protein and that the increase in C3′-endo population does not occur at the expense of the population of C2′-endo conformers. The reduction in bandwidth suggests a possible narrowing in the distribution of C2′-endo-like conformers.

Secondary Structure within the Guanine-Containing Terminal Segments. Examination of the guanine and cytosine base vibrations provides an indication of the helical and base stacking organization of the terminal regions. The band at 680 cm^{-1} is of particular relevance to the discussion of helical structure within the terminal blocks because it is assigned to a conformationally sensitive guanine-coupled furanose vibration. (Most of the guanine residues are in the terminal blocks.) The following correlation between spectral position and configuration has been identified: 680 cm^{-1}, C2′-endo anti; 665

cm^{-1}, C3′-endo anti; 624 cm^{-1}, C3′-endo syn (Nishimura et al., 1983; Benevides & Thomas, 1983). A correlation between 670 cm^{-1} and the C1′-exo and C2′-endo syn configurations has also been suggested (Nishimura et al., 1986). The persistence of the observed 680 cm^{-1} band within the complex indicates that the population of furanose rings in the C2′-endo conformation linked to guanine within the crystalline complex is similar to the population present in solution. An intensity increase is not observed in the region from 660 to 670 cm^{-1}. This suggests that an increase in the C3′-endo population for guanine residues is not observed. Hence, the observed increase in the C3′-endo population identified by analysis of the backbone vibrations does not involve changes in the conformation of furanose rings attached to guanine residues.

The crystalline dodecamer d(CGCGAATTCGCG) contains A-like steps at the site of the type II neokink, that is, CG(3) and CG(9), suggesting that the site of the type II neokink may be a point of accommodation where the dodecamer responds to external forces. Detailed analysis of the Raman spectra of this oligonucleotide in the crystalline and solution states indicates that these A-like steps relax in solution with the oligonucleotide adopting more B-like base stacking patterns (Thomas et al., 1987). This was identified by the following changes proceeding from the crystal to solution: 1255 cm^{-1} (C) decrease and 1299 cm^{-1} (C, A) increase (see Table II and Figure 5). If one compares the spectra of the crystalline dodecamer and the *Eco*RI–oligonucleotide complex, then one observes the following intensity relationships: 1260 cm^{-1} (complex) < 1255 cm^{-1} (dodecamer), 1299 cm^{-1} (complex) > 1299 cm^{-1} (dodecamer), and 1336 cm^{-1} (complex) < 1336 cm^{-1} (dodecamer). The intensity of the band at 1255 cm^{-1} not only is smaller than that observed for the crystalline dodecamer but also is further reduced from that observed for the oligomer in solution. The 1299 cm^{-1} band is greater in the complex than for either the oligonucleotide in solution or the crystalline dodecamer. The changes in the intensities of the bands at 1255 and 1299 cm^{-1} are in the opposite direction to the changes observed going from solution to crystalline forms for the dodecamer d(CGCGAATTCGCG). This suggests that the base stacking in the complex differs from that encountered in solution and in the crystalline dodecamer. The observed changes in base stacking are then not mere extensions of the A-like steps identified in the crystalline dodecamer. The intensity differences in the 1255 and 1299 cm^{-1} bands appear to be linked to changes in base stacking and can be attributed to the protein-induced formation of the type II neokink.

Distortions within the Internal A,T-Containing Segments. Significant helical reorganization appears to occur at the type I neokink which is between the internal GAA and TTC of the recognition sequence. These internal blocks contain all the adenine and most of the thymine residues of the oligonucleotide. Spectral changes in intensity and position which can be assigned to adenine and thymine residues of these blocks are observed in the spectrum of the complex, including intensity decreases at 750 cm^{-1} (T) (shifts to 745 cm^{-1}), 1226 cm^{-1} (T, A), and 1378 cm^{-1} (T, A) as well as intensity increases at 1236 cm^{-1} (T) and 1299 cm^{-1} (A, G). The thymine band at 750 cm^{-1} changes with helical conformation, and recent polarization studies indicate that the band is composed of both in-plane base and out-of-plane base–backbone vibrations (Patapoff et al., 1988). The 1378 cm^{-1} band decreases in intensity and exposes the C, G band at 1367 cm^{-1}. Reductions of the intensity of this band have been observed upon conversion of B DNA to an A- or Z-type helix (Peticolas et al., 1987; Thomas et al., 1987; Wang et al., 1987a,b; Thomas & Peticolas, 1984; Erfurth et al., 1972). The similarity of the changes which accompany the B–A and B–Z transitions arises from the reduction in the degree of base stacking or overlap which is characteristic of both processes. This similarity complicates the formulation of an exact spectroscopic description of the A/T base stacking interactions within the complex. It appears nevertheless that the orientations of the bases within the internal blocks are significantly altered from that observed for B-form DNA.

The spectral changes in the adenine and thymine bands are presumably linked to the formation of the type I neokink. This kink involves unwinding and torsional distortion of the molecule and relative displacements of A–T base pairs (McClarin et al., 1986). In view of the distortions involving the A,T residues and the lack of change in the G residues, it is probable that the increase in C3′-endo population occurs within the internal segments. This does not imply that the internal section adopts an A-DNA-like helical section but that adoption of an C3′-endo conformation may be associated with the formation of the type I neokink.

The adoption of an A-form structure from a B-form structure results in the following frequency changes in adenine and thymine vibrations: 1339 → 1335 cm^{-1} (A); 748 → 745 cm^{-1} (T); 790 → 777 cm^{-1} (T); and 1208 → 1239 cm^{-1} (T) (Thomas et al., 1986; Katahira et al., 1986). The adoption of an A-form helical structure results in a substantial change in base stacking interactions and is accompanied by an increase in the C3′-endo population. Consequently, the changes in frequency described immediately above could be a consequence of changes in either base stacking or furanose conformation or both. If the changes are assumed to be due to prevalence of the C3′-endo conformation and not the A-form DNA stacking interactions, then similar changes might be expected to be observed in the *Eco*RI complex. A thymine band appears at 751 cm^{-1} in the spectrum of the oligonucleotide in solution. The corresponding band appears at 745 cm^{-1} in the crystalline complex. A small displacement in frequency for the band observed at 783 cm^{-1} in solution to 780 cm^{-1} in the complex is also observed. An intensity increase at 1236 cm^{-1} is also observed for the complex. Keeping in mind the above assumptions, these changes are consistent with an increase in the C3′-endo population in the A,T-containing core. Overlap between adenine and guanine contributions in the region about 1338 cm^{-1} prevents identification of a shift in the adenine mode in this region.

Summary and Conclusions

The crystalline complex between the oligonucleotide d(TCGCGAATTCGCG) and the endonuclease *Eco*RI appears to be a product of mutual conformational accommodation between the protein and nucleic acid. Very dramatic changes in the secondary structure of the oligonucleotide are identified by both crystallographic and spectroscopic analysis. The conformational changes spectroscopically identified include (1) an increase in the population of furanose rings adopting the C3′-endo conformation and (2) changes in base stacking which are probably associated with distortions of the phosphate backbone about positions C(3)–G(4) and C(9)–G(10) and unwinding between the symmetry-related segments GAA-TTC. The presence of residual intensity at 851 cm^{-1} in the spectrum of the complex suggests that DNA binding involves a conformational change in the protein. The peak position suggests involvement of a tyrosine residue. Assignment of the residual intensity to protein components due to inadequate subtraction cannot be entirely ruled out, and changes in protein conformation can at best be only tentatively assessed at this

time. Consideration of the crystallographically determined structure of the complex suggests that DNA binding should at the very minimum involve movement of an arm of the protein to encircle the DNA, thereby anchoring the protein to the double helix. Such conformational changes would be consistent with the observed data.

The Raman spectra of the oligonucleotide d-(TCGCGAATTCGCG) in solution and when bound to the protein differ significantly. This is in excellent agreement with the crystallographic results that the neokinks identified in the complex are not observed in the X-ray structure of the nearly identical dodecamer, d(CGCGAATTCGCG). This strengthens the idea that the kinks are a response to the environmental and conformational forces imposed by the protein. This also shows that the neokinks are not preexisting static structural features that can be used by the protein for initial identification of the target restriction sequence. The process of sequence identification therefore involves active deformation of the DNA as the protein pushes and probes the double helix. One interesting possibility is that target sequences are conformationally soft positions where the double helix more readily adopts the deformations imposed by the protein. In this model, base sequence places limitations on the conformational space available to the DNA as well as determining the hydrogen-bonding pattern of recognition. In other words, the actual response of the DNA to the stresses placed upon it by the protein is limited by sequence-specific restrictions on its conformational flexibility.

The situation encountered in the *Eco*RI system suggests that though small topological deformations along the helix may be used for recognition large deformations arise as a secondary response to protein binding. The induction of topological changes in DNA upon protein binding has also been observed for the binding of the ara C protein as well as λ repressors to DNA (Dukunn et al., 1984; Hochshild & Ptashne, 1986). The cooperative binding of two repressor molecules to DNA results in the formation of a DNA loop which allows the two repressors to come into contact with one another. The ability to form a loop is dependent upon the sequence of the DNA between the two binding sections, but the formation of the loop is initiated by protein binding.

Acknowledgments

We thank Drs. Y. Wang and R. Brennan for many hours of insightful and helpful discussion.

Registry No. d(TCGCGAATTCGCG), 90327-46-1.

References

Anderson, W. F., Ohlendorf, D. H., Takeda, Y., & Matthews, B. W. (1981) *Nature 290*, 754.

Anderson, W. F., Takeda, Y., Ohlendorf, D. H., & Matthews, B. W. (1982) *J. Mol. Biol. 159*, 745.

Anderson, J. E., Ptashne, M., & Harrison, S. C. (1985) *Nature 318*, 596.

Benevides, J. M., & Thomas, G. J. (1983) *Nucleic Acids Res. 11*, 5747.

Benevides, J. M., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., & Thomas, G. J., Jr. (1988) *Biochemistry 27*, 931.

Brown, K. B., Kiser, E. J., & Peticolas, W. L. (1972) *Biopolymers 11*, 1838.

Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761.

Drew, H. R., Wing, R. M., Takano, T., Broka, G., Tanaka, S., Itakura, K., & Dickerson, R. E. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2179.

Dukunn, T. M., Hahn, S., Ogden, S., & Schleif, R. F. (1984) *Proc. Natl. Acad. Sci. U.S.A 81*, 5017.

Erfurth, S. C., Kiser, E. J., & Peticolas, W. L. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 938.

Erfurth, S. C., Bond, P. J., & Peticolas, W. L. (1975) *Biopolymers 14*, 1245.

Frederick, C. A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobson, L., Wang, B.-C., Greene, P. J., Boyer, H. W., & Rosenburg, J. M. (1984) *Nature 309*, 327.

Goodman, D. D., & Brahms, J. (1978) *Nucleic Acids Res. 5*, 835.

Greene, P. J., Gupta, M., Boyer, H. W., Brown, E. W., & Rosenberg, J. M. (1981) *J. Biol. Chem. 258*, 2143.

Halford, S. E., & Johnson, N. P. (1980) *Biochem. J. 191*, 593.

Hochschild, A., & Ptashne, M. (1986) *Cell 44*, 681.

Incardonia, N. L., Prescott, B., Sargent, D., Lamda, O. P., & Thomas, G. J. (1987) *Biochemistry 26*, 1532.

Jack, W. E., Rubin, R. A., Newmann, A., & Modrich, P. (1981) in *Gene Amplification and Analysis: Restriction Endonuclease* (Chirikjian, J. G., Ed.) Vol. 1, p 165, Elsevier/North-Holland, Amsterdam.

Jen-Jacobson, L, Kurpiewski, M., Lesser, D., Grable, J., Boyer, H. W., Rosenberg, J. M., & Greene, P. J. (1983) *J. Biol. Chem. 258*, 14638.

Jen-Jacobson, L., Lesser, D., & Kurpiewski, M. (1986) *Cell 45*, 619.

Katahira, M., Nishimura, Y., Sato, T., Mitsui, Y., & Irtaka, Y. (1986) *Biochim. Biophys. Acta 867*, 256.

Kubasek, W. L., Wang, Y., & Thomas, G. A., Patapoff, T. W., Schoenwaelder, K. H., Van der Sande, J. H., & Peticolas, W. L. (1986) *Biochemistry 25*, 7440.

Levitt, M., & Warshel, A. (1978) *J. Am. Chem. Soc. 100*, 2607.

McClarin, J. A., Frederick, C. A., Wang, B.-C., Greene, P. J., Boyer, H. W., Grable, J., & Rosenberg, J. M. (1986) *Science 243*, 1526.

McKay, D. B., & Steitz, T. A. (1981) *Nature 290*, 744.

Modrich, P. (1979) *Q. Rev. Biophys. 12*, 315.

Modrich, P., & Zabel, D. (1976) *J. Biol. Chem. 251*, 5866.

Newmann, A. K., Rubin, R. A., Kim, S.-H., & Modrich, P. (1981) *J. Biol. Chem. 256*, 2143.

Nishimura, Y., Tsuboi, M., Nakano, T., Higuchi, S., Sato, T., Shida, T., Uesugi, S., Ohtsaka, E., & Ikehara, M. (1983) *Nucleic Acids Res. 11*, 1379.

Nishimura, Y., Tsuboi, M., Sato, T., & Aoki, K. (1986) *J. Mol. Struct. 146*, 123.

Olsen, W. K., & Sussman, J. L. (1982) *J. Am. Chem. Soc. 104*, 270.

Pabo, C. O., & Lewis, M. (1982) *Nature 298*, 443.

Patapoff, T. W., Thomas, G. A., & Peticolas, W. L. (1988) *Biopolymers 27*, 493.

Peticolas, W. L., Kubasek, W. L., Thomas, G. A., & Tsuboi, M. (1987) in *Biological Applications of Raman Spectroscopy* (Spiro, T., Ed.) Vol. 1, p 81, John Wiley and Sons Inc., New York.

Rosenberg, J. M., Boyer, H. W., & Greene, P. J. (1981) in *Gene Amplification and Analysis: Restriction Endonuclease* (Chirikjian, J. G., Ed.) Vol. 1, p 131, Elsevier/North-Holland, Amsterdam.

Schevitz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L., & Sigler, P. B. (1985) *Nature 317*, 782.

Shaked, Z., Rabinovich, D., Kennard, O., Cruse, W. B. T., Salisbury, A., & Viswamitra, M. A. (1983) *J. Mol. Biol. 166*, 183.

Siamwiza, M. N., Lord, R. C., Chen, M. C., Takamatsu, T., Harada, I., & Shimanouchi, T. (1975) *Biochemistry 14*, 4870.
Thomas, G. A., & Peticolas, W. L. (1983) *J. Am. Chem. Soc. 105*, 993.
Thomas, G. A., & Peticolas, W. L. (1984) *Biochemistry 23*, 3202.
Thomas, G. A., Patapoff, T. W., Wang, Y., & and Peticolas, W. L. (1987) in *Conservation in Biomolecular Stereodynamics, Book of Abstracts* (Sarma, R. H., Ed.) pp 32–33, State University of New York at Albany, Albany, NY.
Thomas, G. J. (1987) in *Biological Applications of Raman Spectroscopy* (Spiro T., Ed.), Vol. 1, p 135, John Wiley and Sons Inc., New York.
Thomas, G. J., & Hartman, K. H. (1973) *Biochim. Biophys. Acta 312*, 3111.
Thomas, G. J., Benevides, J. M., & Prescott, B. (1986) in *Biomolecular Stereodynamics IV* (Sarma, R. H., & Sarma, M. N., Eds.) p 227, Adenine Press, Albany, NY.
Wang, Y., Thomas, G., A., & Peticolas, W. L. (1987a) *Biochemistry 26*, 5178.
Wang, Y., Thomas, G. A., & Peticolas, W. L. (1987b) *J. Biomol. Struct. Dyn. 5*, 249.
Wing, R., Drew, H., Takano, T., Broka, C., Tanka, S., Itakura, K., & Dickerson, R. E. (1980) *Nature 287*, 755.
Wartell, R. M., & Harrel, J. T. (1986) *Biochemistry 25*, 2664.

Localized Chemical Hyperreactivity in Supercoiled DNA: Evidence for Base Unpairing in Sequences That Induce Low-Salt Cruciform Extrusion†

Judy C. Furlong, Karen M. Sullivan,‡ Alastair I. H. Murchie, Gerald W. Gough,§ and David M. J. Lilley*
Department of Biochemistry, The University, Dundee DD1 4HN, U.K.
Received April 20, 1988; Revised Manuscript Received October 20, 1988

ABSTRACT: Certain A+T-rich DNA sequences (C-type inducing sequences) cause adjacent inverted repeats to undergo cruciform extrusion by a particular pathway (C-type extrusion), which is characterized by large activation energies and extrusion at low salt concentrations and relatively low temperatures. When they are supercoiled, these sequences become reactive toward the normally single-strand-selective reagents bromoacetaldehyde, glyoxal, osmium tetraoxide, and sodium bisulfite. The following evidence is presented: (1) The most reactive sequences are those to the left of the inverted repeat. (2) Chemical reactivity is suppressed by either sodium chloride or micromolar concentrations of distamycin. The suppression of reactivity closely parallels that of C-type cruciform extrusion. (3) Chemical reactivity requires a threshold level of negative supercoiling. The threshold superhelix density depends on the prevailing salt concentration. (4) Analysis of temperature dependences suggests that reaction with osmium tetraoxide involves transient unstacking events, while bromoacetaldehyde requires larger scale helix opening. Thus a variety of opening events may occur in the supercoiled A+T-rich sequences, from small-amplitude breathing to low-frequency, large-amplitude openings. The latter appear to be responsible for C-type cruciform extrusion.

Contextual effects on cruciform extrusion provide a striking example of the transmission of structural effects in supercoiled DNA (Sullivan & Lilley, 1986). C-type inducing sequences are very A+T-rich regions of DNA which act upon nearby inverted repeats, changing the entire character of the process by which the molecule isomerizes to form a cruciform structure. All available evidence (Lilley, 1985; Sullivan & Lilley, 1986, 1987, 1988; Sullivan et al., 1988) suggests that the inducing sequences are responsible for a change in the mechanistic pathway for the extrusion process. The C-type pathway is beleved to proceed via a transition state which is well represented by a large melted "bubble" in the DNA, which may then undergo intrastrand base pairing to form a fully extruded cruciform in a single step. It is characterized kinetically by a very large activation energy (in the region of 200 kcal mol^{-1}) and extrusion in the absence of added salts. The C-type mechanism involves large-scale unpairing of bases in the DNA. The properties of the inducing sequences may be summarized as follows:

(1) The sequences are very A+T rich (≥70%). Application of statistical thermodynamic DNA melting theory (Schaeffer et al., 1988) indicates that the inducing sequences have a high propensity to cooperative melting. This is further substantiated by the observation that sequences of normal base sequence become quasi-C-type in character if helical stability is reduced by addition of solvents such as formamide (Sullivan & Lilley, 1988).

(2) Inducing sequences operate on neighboring inverted repeats in cis, without regard to polarity (Sullivan & Lilley, 1986). The nature of sequences (termed transmitting sequences) in between the inducing sequence and the inverted repeat is important (Sullivan et al., 1988).

(3) A single C-type inducing sequence is sufficient to confer C-type character on a nearby inverted repeat in 0 mM sodium chloride (Sullivan & Lilley, 1986).

(4) The effect of a C-type inducing sequence may be modulated by changes in sequence at least 100 bp distant from

† We thank the Cancer Research Campaign and the Royal Society for financial support.
* To whom correspondence should be addressed.
‡ Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge, U.K.
§ Present address: ICRF Clare Hall Labs, South Mimms, London, U.K.

0006-2960/89/0428-2009$01.50/0