Protein-Directed DNA Structure II. Raman Spectroscopy of a Leucine Zipper bZIP Complex†

James M. Benevides,‡ Tiansheng Li,§ Xiang-Jun Lu,|| A. R. Srinivasan,|| Wilma K. Olson,|| Michael A. Weiss,*,⊥ and George J. Thomas, Jr*,‡

Division of Cell Biology and Biophysics, School of Biological Sciences, University of Missouri−Kansas City, Kansas City, MO 64110, Department of Chemistry, Rutgers University, New Brunswick, New Jersey 08903, and Center for Molecular Oncology, Department of Biochemistry and Molecular Biology and Department of Chemistry, University of Chicago, Chicago, Illinois 60637

Received January 8, 1999; Revised Manuscript Received October 29, 1999

ABSTRACT: Mechanisms of transcription may involve protein-directed changes in DNA structure and DNA-directed changes in protein structure. We have employed Raman spectroscopy to characterize vibrational signatures associated with such induced molecular fitting for two classes of transcription factors—the basic leucine-zipper (bZIP) motif and the high-mobility-group (HMG) box—each with a DNA target site. Results for bZIP are described here; findings for the HMG-box are reported in the preceding paper in this issue [Benevides, J. M., Chan, G., Lu, X.-J., Olson, W. K., Weiss, M. A., and Thomas, G. J., Jr. (2000) *Biochemistry 39*, 537−547]. The yeast activator GCN4 provides a well-studied example of bZIP recognition, wherein B-DNA serves essentially as a template for protein folding. Analysis of Raman spectra of the 57-residue GCN4 bZIP domain, its AP-1 binding site, and their specific complex confirms a DNA-induced increase in α-helicity, attributable to folding of GCN4 basic arms with virtually no change in B-DNA structure, consistent with previous X-ray and NMR structure determinations. The absence of DNA perturbations in the bZIP model contrasts sharply with the HMG box, where DNA structure perturbations predominate. The bZIP and HMG-box models represent two opposing extremes in a range of induced fits identifiable by Raman spectroscopy. Previously characterized λ repressor/operator complexes [Benevides, J. M., Weiss, M. A., and Thomas, G. J. (1994) *J. Biol. Chem. 269*, 10869−10878] occupy an intermediate position within this range. A comprehensive tabulation of Raman markers proposed as diagnostic of different protein/DNA recognition motifs is presented. The results are analyzed in terms of available DNA crystal structures (Nucleic Acid Database) to identify details of DNA conformation that correlate with specific Raman recognition markers.

Sequence-specific recognition of DNA by gene-regulatory proteins may involve the induced fit of either or both interacting macromolecules (*1*). The extent of induced fit spans a range that is determined by biological function. At one extreme are large-scale conformational changes in the DNA site with relatively little change in the structure of the protein. For example, minor-groove-binding proteins containing the high-mobility-group (HMG)[1] box (*2*, *3*) have the capacity to induce major structural reorganizations of their respective DNA-binding sites. In such recognition, proteins containing well-defined elements of secondary structure and overall tertiary fold may undergo enhanced ordering upon DNA binding. Conversely, the Fos-Jun heterodimer (*4*) and yeast transcriptional activator GCN4 (*5*−*8*), both of which are basic leucine-zipper (bZIP) proteins, undergo dramatic disorder−order transitions upon binding to DNA without inducing appreciable change in the conformation of the B-DNA template. Intermediate in this scheme are major-groove binding proteins that utilize the helix-turn-helix (HTH) motif, exemplified in eukaryotes by the homeodomain (*9*) and in prokaryotes by the cI repressor of phage λ (*10*). In these systems, modest rearrangements of protein structure and localized changes in DNA structure occur.

In principle, a detailed structural characterization of induced fit requires independent analysis of a range of protein/DNA complexes and their isolated components by high-resolution methods. In practice, complete characterizations have been accomplished in only three cases, viz., X-ray

† Paper LXXIII in the series Raman Spectral Studies of Nucleic Acids. Supported by NIH Grants GM54378 (G.J.T.), HD33462 (M.A.W.), and GM20861 (W.K.O.). M.A.W. was an American Heart Association Established Investigator and a Lucille Markey Scholar, University of Chicago.

* To whom correspondence may be addressed. (G.J.T.) Phone: (816) 235-5247. E-mail: thomasgj@umkc.edu. (M.A.W.) Phone: (216) 368-5991. E-mail: weiss@biochemistry.cwru.edu.

‡ University of Missouri−Kansas City.

§ Present address: Department of Pharmaceutics, Amgen Inc., Thousand Oaks, CA 91320.

|| Rutgers University.

⊥ University of Chicago. Present address: Department of Biochemistry, Wood W427, Case Western Reserve School of Medicine, 10900 Euclid Avenue, Cleveland, OH 44106-4935.

[1] Abbreviations: bp, base pair; bZIP, basic region of the leucine-zipper motif; HMG, high-mobility-group; hSRY, sex-determining region of the human Y chromosome; hSRY−HMG box, high-mobility-group box encoded by the human sex-determining region of the Y chromosome; HTH, helix-turn-helix; NMR, nuclear magnetic resonance; TBP, TATA-box binding protein; UVRR, ultraviolet-resonance Raman spectroscopy.

10.1021/bi990053x CCC: $19.00
Published on Web 12/23/1999

structures of the *trp* operator (*11*), *EcoR1* sites (*12*), and E2-DNA targets (*13*), with and without their bound proteins. Barriers are posed by the difficulty of obtaining well-ordered crystals of B-DNA, issues of structural interpretation for the crystal lattice, and a continuing debate regarding the accuracy and precision of NMR structures of nucleic acids (*14*, *15*). We have employed Raman difference spectroscopy as an alternative, wherein the spectrum of the complex is compared with the sum of spectra of the isolated components. This approach is particularly well suited to analysis of protein/DNA recognition in solution. It enables systematic comparison of the structures of many gene-regulatory complexes, although with limited site-specific resolution (*16*−*20*).

In the preceding paper in this issue (*20*), we describe the application of Raman spectroscopy to the HMG-box of a putative transcription factor (hSRY) encoded by the human sex-determining region of the Y chromosome and its specific complex with a DNA target site, d(GAACAATC)·d(GATTGTTC). The hSRY−HMG:DNA complex is formed with minimal change in Raman markers of the hSRY−HMG box, but with extraordinarily large perturbations to Raman markers associated with DNA backbone geometry and base stacking. These serve as the Raman signature of minor-groove recognition and are proposed as diagnostic of sharp bending of the B-DNA site. They establish one end of the range of induced fit. The opposite end of the induced-fit spectrum is represented here by the specific complex between the DNA-binding domain of the yeast GCN4 bZIP protein and its pentadecanucleotide binding site (AP-1). The GCN4:DNA complex, which is the focus of the present study, consists of the divergent coiled coil of α-helices shown in Figure 1A. At physiological conditions and in the absence of the DNA template, the bZIP motif is a random coil. DNA binding induces dimerization and a concurrent α-helical folding transition in the basic arms of each subunit; the protein-bound DNA remains a double helix of the canonical B-form. Investigation of the GCN4 domain, AP-1 target site, and their specific complex by Raman spectroscopy demonstrates that formation of the complex at physiological temperature occurs without appreciable conformational change in the DNA site. GCN4:DNA thus serves as the converse of the hSRY−HMG:DNA recognition model described previously (*20*).

Included in the present analysis is a reassessment of the Raman difference spectrum for the DNA-binding domain of the λ repressor in a specific complex with its O_L1 operator site (cI:DNA), the structure of which is shown in Figure 1B. Tabulation of Raman markers of the three models (hSRY−HMG:DNA, cI:DNA, and GCN4:DNA) furnishes an empirical database against which other specific protein/DNA complexes may be compared. Finally, the Raman results are analyzed in terms of experimentally determined DNA structure parameters (Nucleic Acid Database) in order to identify the probable features of DNA conformation that correlate with the observed Raman recognition markers.

MATERIALS AND METHODS

1. Preparation and Purification of the GCN4 DNA-Binding Domain. Procedures for expression and purification of the DNA-binding domain of GCN4 (residues 225−281) have been described (*8*). Briefly, the polypeptide was overexpressed in *Escherichia coli* using the T7 system and purified by a combination of ammonium sulfate precipitation and phosphocellulose chromatography. The 59-residue sequence examined here contains two nonnative residues at the N-terminus (MK, encoded by the expression vector) followed by a 57-residue DNA-binding domain (SDPAA^5 LKRAR^{10} NTEAA^{15} RRSRA^{20} RKLQR^{25} MKQLE^{30} DKVEE^{35} LLSKN^{40} YHLEN^{45} EVARL^{50} KKLVG^{55} ER). The sequence was verified by composition analysis and N-terminal sequencing.

2. Oligonucleotide Synthesis. The pentadecameric AP-1 duplex [d(GAGATGAGTCATCTC) and its complement d(GAGATGACTCATCTC)] were synthesized using automated phosphoramidite chemistry. The AP-1 site resembles the self-complementary hexadecameric duplex bound by the mammalian ATF/CREB family of proteins (*8*).

3. Sample Preparation for Raman Spectroscopy. The purified GCN4 peptide was dissolved to a final concentration of 100 μg/μL in a solution containing 0.1 M KCl + 10 mM Tris at pH 7.0 and a 10 μL aliquot was sealed in a glass capillary (Kimax no. 34507) for Raman data collection. The purified AP-1-binding site was dissolved to a final concentration of 50 μg/μL in the same buffer. GCN4:DNA complexes were prepared at 2:1 and 3:1 molar ratios following described procedures (*8*). For data collection, samples were thermostated to ±1 °C at the temperature indicated.

Raman spectra were excited with the 514.5 nm line of a Coherent Innova 70 argon-ion laser, using approximately 100 mW of radiant power at the sample. Spectra were collected on a model 1877 spectrometer (Spex Industries, Edison, NJ) interfaced with an OMA-III intensified diode array detector (EG&G, Princeton, NJ) as described (*21*). Spectra were averaged from 40 exposures of 30 s each. Additional data were recorded on a Ramalog 1401 scanning spectrometer (Spex Industries) with spectral slit width of 8 cm^{-1} and integration time of 1.5 s/cm^{-1}. Ten spectral scans were averaged for the region 400−1800 cm^{-1}.

4. Conformational Analysis. DNA dimer steps from protein/DNA complexes in the Nucleic Acid Database (*22*) and Protein Data Bank (*44*)−PDR010 [the 1.8 Å resolution structure of the cI λ repressor-operator complex (*10*)], PDT002 [the 2.9 Å resolution structure of GCN4 bound to a 20-bp DNA duplex (*5*) containing the AP-1 binding site examined in this work], and PDT029 [the 2.2 Å resolution structure of GCN4 bound to the symmetric ATF/CREB of DNA (*7*)]—were analyzed at the backbone and base-pair level using the approach described in the companion article (*20*). Dimer step parameters in the three complexes were further compared against the base-sequence dependent mean values $\theta°$ (*23*) and standard deviations $\Delta\theta$ (*24*) collected for high-resolution crystalline B-DNA steps. Parameter deformations provide a measure of distortion of B-DNA in the crystal complex.

RESULTS AND DISCUSSION

GCN4 consists of a C-terminal leucine-zipper region that functions as the interface for subunit dimerization and an N-terminal basic region that binds to the major groove of B-DNA. Dimerization and folding are coupled to specific DNA binding (*4*−*8*), as revealed in the crystal structure of the GCN4:DNA complex (Figure 1A) (*5*). At the solution

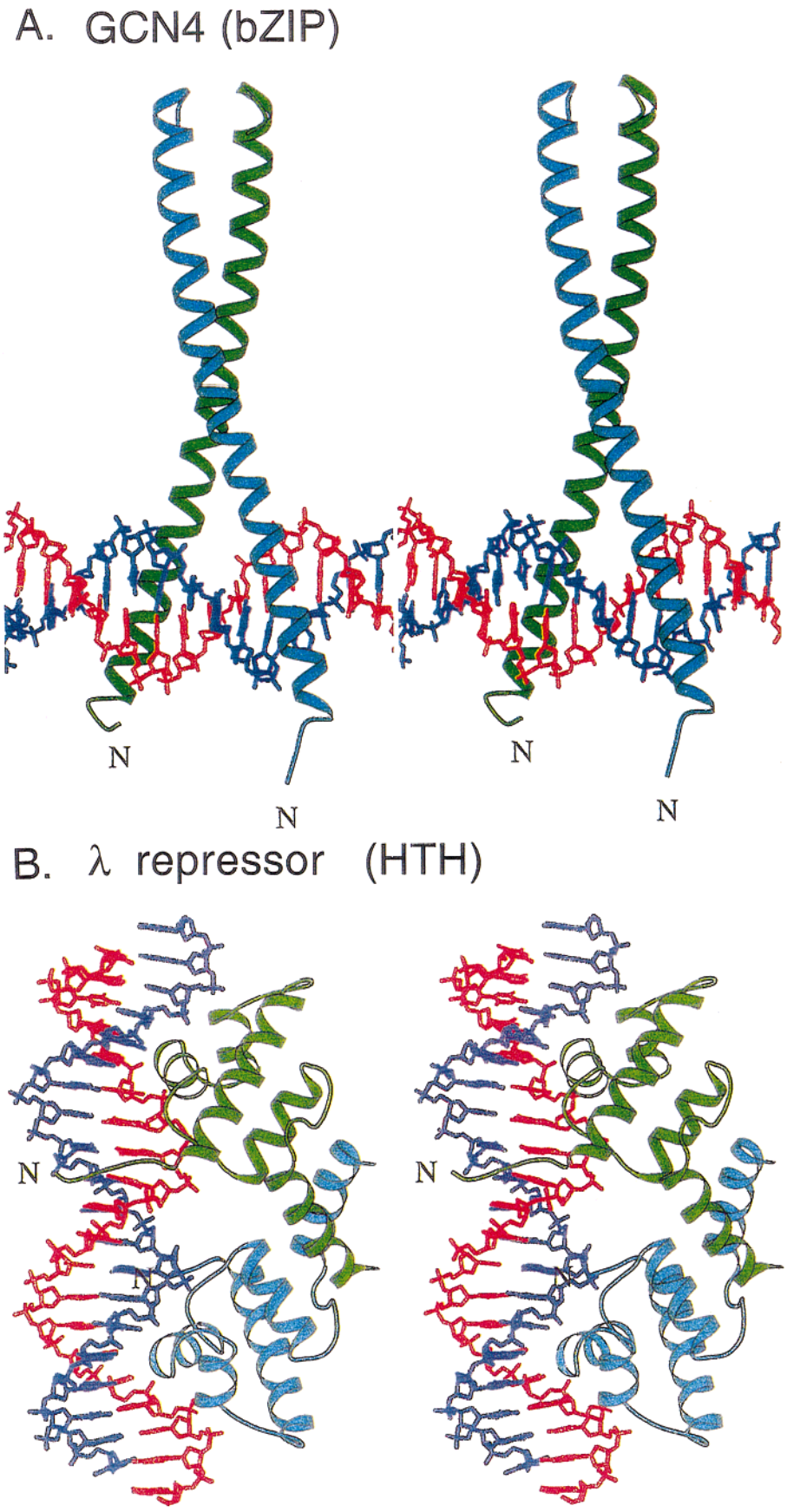

FIGURE 1: (A) Stereo image of the crystal structure of the complex formed between the coiled-coil leucine zipper (bZIP) domain of the yeast transcriptional activator GCN4 and its AP-1-binding site. (B) Stereo image of the crystal structure of the complex formed between the DNA-binding domain of the phage λ cI repressor and its O_L1 operator site (*10*). Emulations of DNA structures in the complexes were facilitated by use of images obtained through the courtesy of Drs. T. Chiu and R. E. Dickerson, UCLA (*41*).

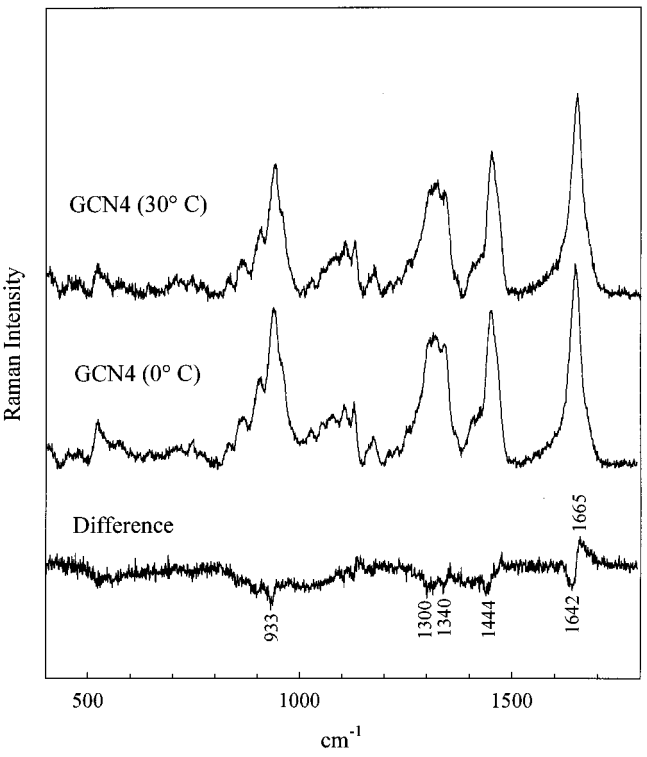

FIGURE 2: Raman spectra (400−1800 cm^{-1}) of the 59-residue DNA-binding domain of the yeast transcriptional activator GCN4 at 100 $\mu g/\mu L$ in 100 mM KCl and 10 mM Tris, pH 7.0. (Top) 30 °C spectrum. (Middle) 0 °C spectrum. (Bottom) Difference spectrum (30 minus 0 °C). Difference troughs at 1444 cm^{-1} and at 933, 1300, 1340, and 1642 cm^{-1} indicate, respectively, altered aliphatic side-chain environments and diminished α-helix at 30 °C.

concentrations employed for Raman spectroscopy, the protein exhibits appreciable α-helical secondary structure, presumably in the form of a partially folded dimer containing an intact coiled coil. The structure of the isolated coiled coil has been determined at high resolution (*25*) and is similar to that observed in the crystal structure of the DNA complex. Although folding of nascent helices in the basic region occurs upon DNA binding (*4*, *8*), the structure of the isolated domain has not been determined by either NMR spectroscopy or X-ray crystallography. The present study is designed to test the hypothesis that the AP-1 site provides a preformed DNA template for protein folding. This hypothesis, which stands in contrast to previous Raman studies of protein-induced changes in DNA structure by the cI and Cro repressors of phage λ (*16*−*18*, *26*), is motivated by detailed analysis of the crystal structure of the GCN4:DNA complex (*5*). As described below, Raman spectra show that the DNA component of this complex exhibits structural parameters, including both backbone torsion angles and base-pair geometries, that are characteristic of B-DNA. Because X-ray, NMR, and Raman studies of many other specific protein/DNA complexes show significant deviations of DNA from the canonical B-form (*16*−*20*), the absence of such deviations in the GCN4:DNA complex is noteworthy. Raman studies of the isolated protein, isolated DNA, and their specific complex are presented below. An analysis of the thermal stability and unfolding of the complex is also given.

1. Secondary Structure of GCN4. The Raman spectrum of the free protein (GCN4) in the region 400−1800 cm^{-1} is shown in Figure 2. The spectrum at 0 °C (Figure 2, middle trace) displays intense Raman bands diagnostic of α-helix at 1646 cm^{-1} (amide I), 1343 (main chain C−C stretch), 1300 cm^{-1} (amide III), and 933 cm^{-1} (side-chain C−C stretch) (*16*−*18*, *27*−*30*), in accord with previous NMR (*8*) and X-ray studies (*5*). Qualitative assessment of secondary structure by Fourier deconvolution of the amide I band profile (Figure 3, right) suggests nominal amounts of irregular structures and turns (1660 and 1673 cm^{-1} shoulders) in addition to the predominant α-helix (1646 cm^{-1}). Similarly, Fourier deconvolution of the 1250−1375 cm^{-1} region indicates amide III bands at 1250 (irregular/turns) and 1282/1298 cm^{-1} (α-helix), in addition to side-chain markers: Tyrosine generates the 1263 cm^{-1} feature; deformation modes of aliphatic side chains are the major contributors to bands at 1321 and 1365 cm^{-1}; and the α-helical main chain is a major contributor to the very intense band at 1343 cm^{-1} (*29*, *30*). Relative intensities of deconvolved amide I and amide III bands imply 80−90% α-helix and 10−20% irregular structure in GCN4 at the experimental conditions employed, in agreement with the ∼80% α-helix estimated from CD measurements under similar conditions (*8*).

CD studies have shown that the α-helicity of GCN4 decreases from ∼80% at 10 °C to ∼70% at 25 °C (*8*, *31*). To further assess the thermal unfolding process, including

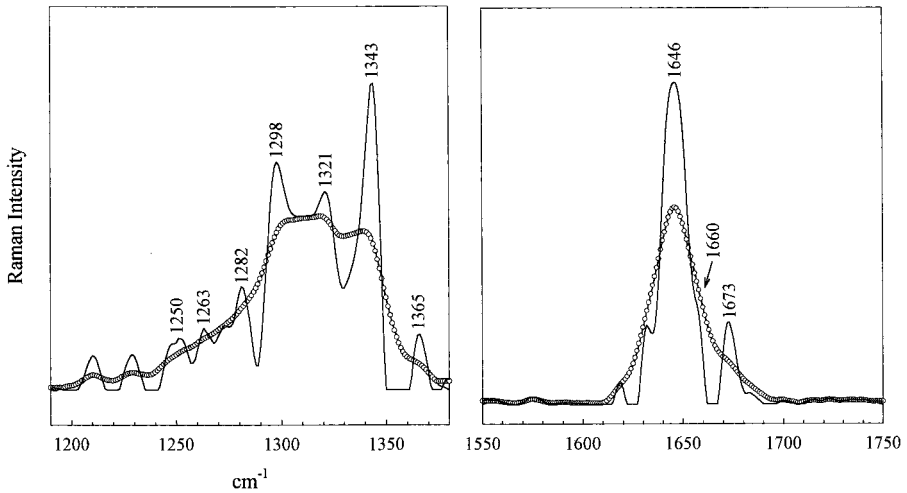

FIGURE 3: Fourier deconvolutions (- - -) of experimental (○) Raman amide III (left panel) and amide I (right panel) profiles of the GCN4:DNA-binding domain. Labels indicate wavenumber values of deconvolved band components. See discussion in text and ref *42*.

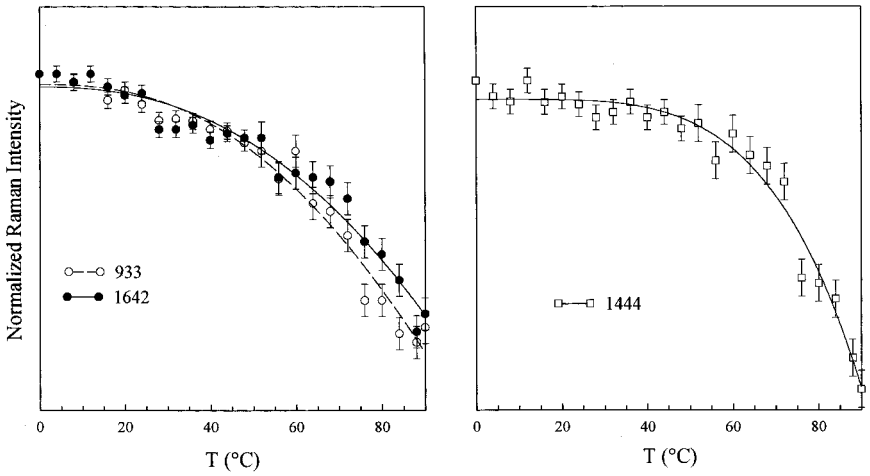

FIGURE 4: Temperature-dependent intensities of Raman bands diagnostic of GCN4 α-helix [left panel, 933 cm^{-1} (○) and 1642 (●) cm^{-1}] and aliphatic side-chain environment [right panel, 1444 cm^{-1} (□)].

roles of both the peptide main chain and specific side chains, the Raman signature of GCN4 has been characterized in the range 0−90 °C. Spectra obtained at 30 (top trace) and 0 °C (middle trace) are compared in Figure 2. The computed difference spectrum (Figure 2, bottom trace) demonstrates diminished intensities of all α-helix markers at the higher temperature, evidenced by difference troughs at 1642, 1300, 1340, and 933 cm^{-1}. Accompanying this loss of α-helix is a corresponding increase in irregular and/or β-strand structure, evidenced by a difference peak at 1665 cm^{-1}. The magnitude of the amide I intensity shift implies a 16 ± 4% loss of helix between 0 and 30 °C, consistent with CD results (*8*, *31*). The difference spectrum of Figure 2 also exhibits a trough at 1444 cm^{-1} (methylene marker), corresponding to a net change in environment of aliphatic side chains between 0 and 30 °C.

Normalized intensities of key Raman markers are plotted vs temperature (0−90 °C range) in Figure 4. Indicators of α-helix (933 and 1646 cm^{-1}, left panel) reveal a dramatic loss of secondary structure with increasing temperature. In addition to the previously noted attenuation of α-helicity between 0 and 30 °C (compatible with a helix-coil transition of the N-terminal arm of GCN4), more extensive helix melting is observed between 50 and 90 °C, presumably corresponding to cooperative unfolding and dissociation of the C-terminal coiled coil (*8*, *31*).

Melting of the coiled coil involves unfolding of specific side-chain/side-chain interactions (leucine zipper) (*25*). The temperature-dependent profile of the methylene marker at 1444 cm^{-1} (Figure 4, right panel) reveals a large intensity change corresponding to thermal unfolding of the coiled-coil (50−90 °C), but minimal intensity change corresponding to thermal unfolding of the basic N-terminal domain (0−30 °C). Because methylene-containing side chains contributing to the 1444 cm^{-1} marker (*30*) are distributed throughout the protein sequence, the data of Figure 4 (right panel) imply that the observed temperature dependence is diagnostic of the unfolding/dissociation transition of the leucine zipper. We presume that side-chain conformations in the divergent α-helical arms are not well ordered even at low temperature due to the absence of a helix−helix interface.

2. Secondary Structure of AP-1 DNA. Assignments of bands in the Raman spectrum of AP-1 (Figure 5) are based upon previous studies of B-DNA and have been discussed in detail (*32*, *33*). Diminished hypochromism at 30 °C, due to partial unstacking of bases at the ends of the duplex, accounts for difference peaks at 668 (G, T), 725 (A), 779 (C), 1330 (A, G), 1368 (G), 1483 (G, A), and 1572 (A, G) cm^{-1} (*33*−*35*). Nevertheless, Raman markers of the deoxyribose-phosphate moieties (800−900 cm^{-1}) are not altered appreciably between 0 and 30 °C, indicating that B-form

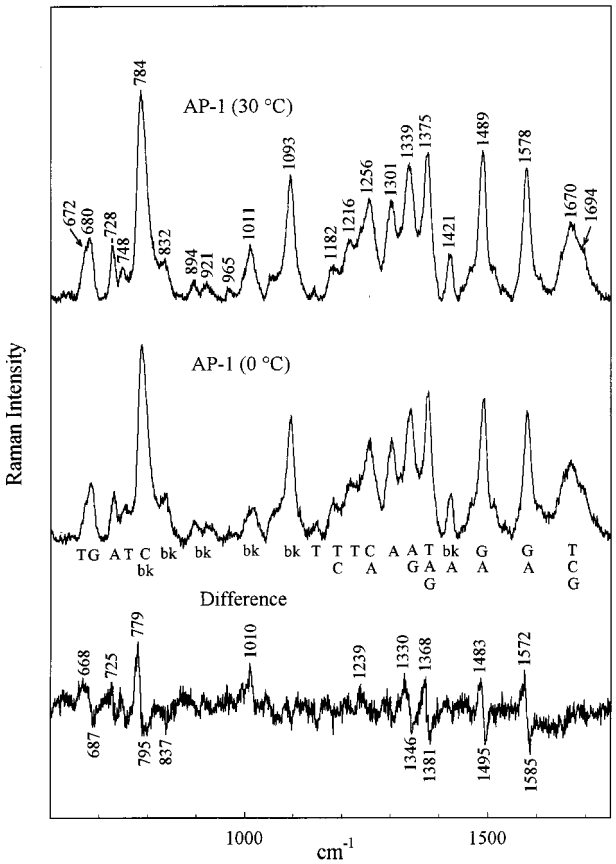

FIGURE 5: Raman spectra (600−1750 cm^{-1}) of the pentadecanucleotide AP-1 duplex, d(GAGATGAGTCATCTC)·d(GAGATGACTCATCTC) at 30 $\mu g/\mu L$ in 100 mM KCl, 10 mM Tris, pH 7.0. (Top) 30 °C spectrum. (Middle) 0 °C spectrum. (Bottom) Difference spectrum (30 minus 0 °C). Assignments of major bands to base (A, C, G, T) and backbone (bk) residues are indicated for the 0 °C spectrum. Difference peaks and troughs reflect partial melting (fraying) between 0 and 30 °C.

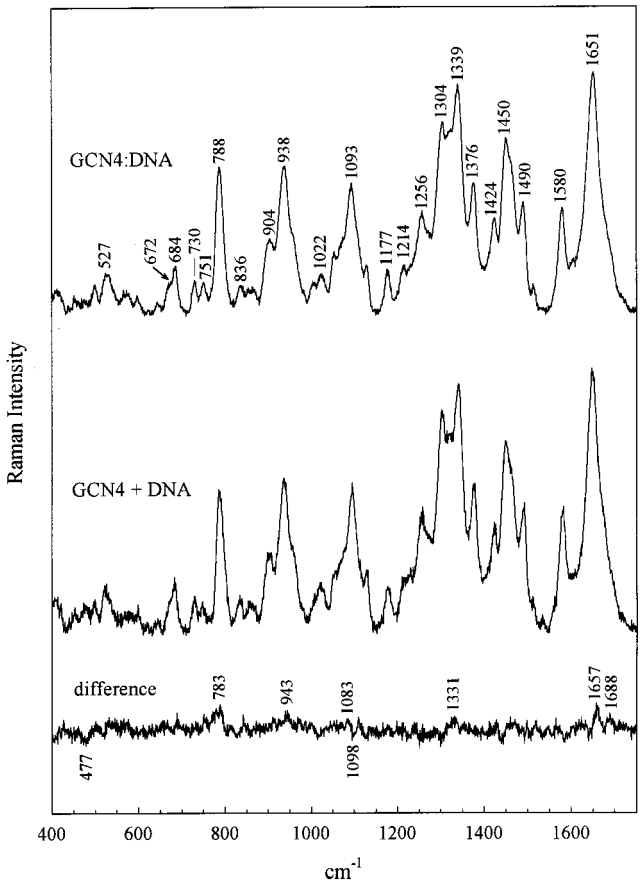

FIGURE 6: Raman spectrum of the specific GCN4:DNA complex (top trace) and the spectral sum of constituents (middle trace), obtained from samples at 0 °C. The difference spectrum (bottom trace) reveals only small spectral perturbations attendant with complex formation.

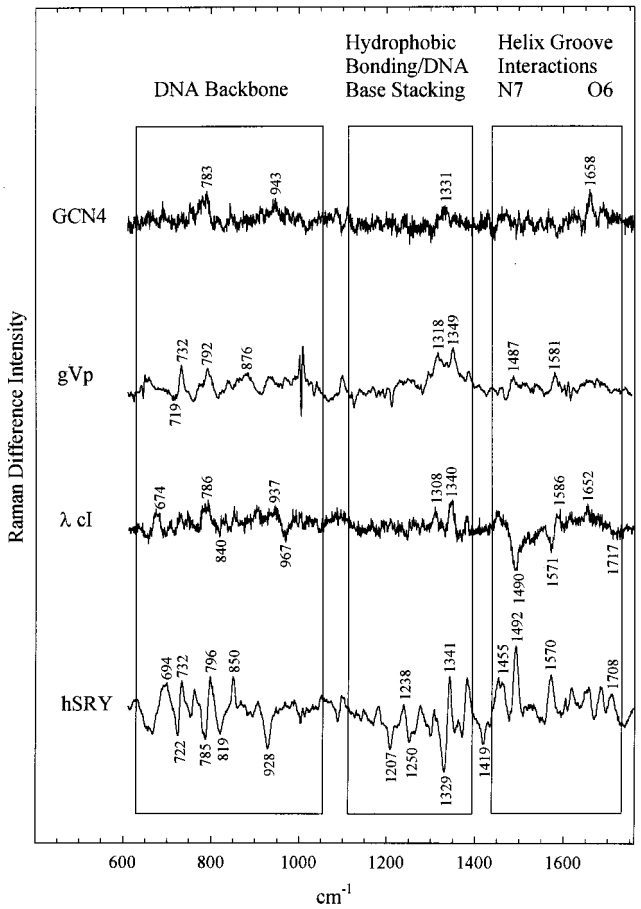

FIGURE 7: Raman difference spectra reflecting DNA reorganizations due to different DNA-binding proteins. From top to bottom: bZIP protein GCN4 binding to an AP-1 target site; ssDNA-binding protein of M13 phage (gVp) binding to the ssDNA analogue, poly(dA); cI repressor of phage λ binding to its O_L1 target site; human sex determining factor hSRY−HMG box binding to its DNA target site. The Raman signature is minimally perturbed in the unbent GCN4:DNA complex and maximally perturbed in the sharply bent hSRY−HMG:DNA complex (*20*). Ordinates of the difference spectra have been scaled to reflect the same 1090 cm^{-1} intensity in the parent complex. Labeling of boxed segments refers to bands of DNA.

backbone geometry is largely retained up to 30 °C. ^{1}H NMR studies of this DNA site (data not shown) lead to the same conclusion.

3. Raman Spectroscopy of the Complex between GCN4 and AP-1. (a) Overview. The Raman spectrum of the specific GCN4:DNA complex at 0 °C (2 mol of GCN4 59-mer/mol of pentadecanucleotide AP-1 duplex) is shown in the top trace of Figure 6. The spectral sum of constituents at the same experimental conditions is shown in the middle trace. The corresponding difference spectrum (Figure 6, bottom) reveals no substantial changes in Raman bands of protein or DNA. Possible difference peaks occur invariably at the positions of the strongest bands in the parent spectra, namely near 783, 943, 1331, and 1657 cm^{-1}, and do not meet the criteria of significance (*20*). Accordingly, any relative intensity changes are marginal. Such very weak features imply minimal perturbations to the α-helix of GCN4 at low temperature (*8*) and essentially no perturbation of DNA. The overall B-conformation of the AP-1 site is therefore not affected significantly by GCN4 binding. The altogether featureless difference spectrum of Figure 6 is unusual and distinctive; it contrasts sharply with the prolific Raman difference signature associated with the hSRY−HMG:DNA complex (*20*).

Comparison of Raman signatures of the GCN4:DNA complex and sum of constituents at 30 °C (data not shown) indicates that the complex is largely intact at 30 °C and that fraying in protein-free AP-1 (Figure 5) is attenuated by bound GCN4. Protein damping of B-DNA dynamics has also been reported for other DNA:protein complexes (*36*). α-Helix markers are more prominent for the complex at 30 °C (data not shown) than for the free protein at 0 °C, consistent with an increase in dimerization and α-helical folding induced by DNA binding at 30 °C.

(b) Interactions of DNA Major-Groove Sites. The GCN4:DNA crystal structure reveals several contacts involving GCN4 side chains with either hydrogen-bond acceptor groups of the DNA bases or thymine $C5H_3$ sites (*5*). The Raman spectral interval 1350−1750 cm^{-1} contains bands that are, in principle, sensitive to such specific protein/DNA interactions (*16*−*20*, *37*). For example, a guanine band shifts from 1490 to 1470 cm^{-1} with N7 hydrogen bonding, and a thymine marker near 1376−1380 cm^{-1} undergoes intensity increase with increasing hydrophobic contact. Both perturbations are observed in the λ cI:DNA complex (*18*). In the case of the GCN4:DNA complex, only one guanine N7 contact is expected in the pentadecameric DNA site, and thus the anticipated spectral perturbation is not of sufficient magnitude to be observed over the noise level of the Figure

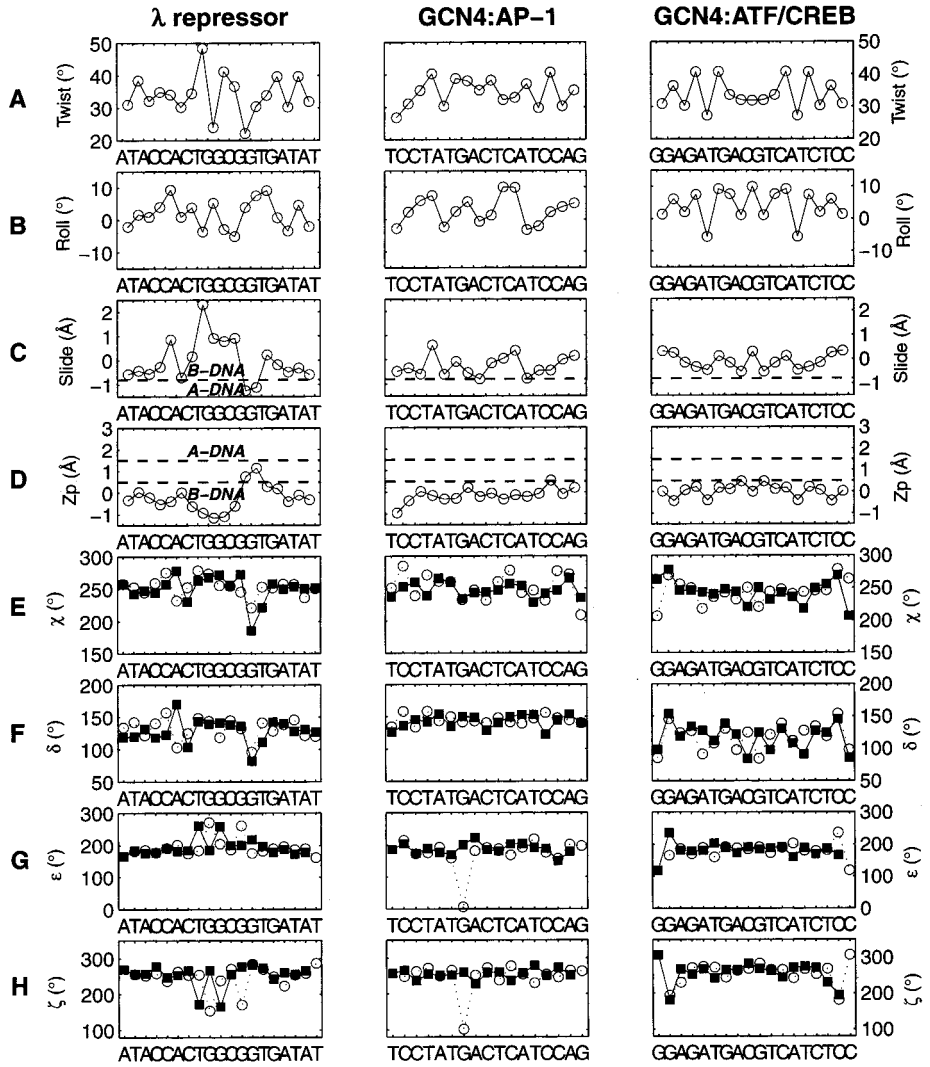

FIGURE 8: Comparison of X-ray determined structural parameters of DNA sequences in complexes with major-groove-binding proteins (*5*, *10*): λ repressor cI (left column), GCN4 (middle and right columns). The respective panels (rows A−H, from top to bottom) show sequence-related variations in base twist angle, base roll angle, base slide (Å), Z_P (Å), χ torsion, δ torsion, ϵ torsion, and ζ torsion, as described previously (*20*). Solid squares and open circles in rows E−H distinguish torsions along the sequence strand and complementary strand, respectively. DNA from the NMR ensembles is described in terms of the averages and standard deviations (error bars) of parameters at individual base-pair steps. See also ref *20*.

6 data. Similarly, thymine $C5H_3$ sites of GCN4:DNA are not sufficiently perturbed to generate a discernible difference feature near 1376−1380 cm^{-1}. Figure 7 compares Raman difference spectra for several models of protein/DNA recognition and demonstrates the dramatic differences between the GCN4:DNA and cI:DNA complexes.

4. Conformational Analysis of Major-Groove Recognition. In the preceding paper in this issue (*20*), we examined possible relationships between DNA Raman bands perturbed in the hSRY−HMG:DNA complex and the specific DNA helical parameters that are altered by minor-groove binding proteins. A similar analysis is given here (Figure 8) for the major-groove binding mechanisms of HTH [λ cI repressor-operator (*10*)] and bZIP [GCN4:AP-1 (*5*) and GCN4:ATF/CREB (*7*)].

Notwithstanding different resolutions in crystal and NMR structures (see above), Figure 8 shows that the λ operator DNA is more severely deformed by protein than either of the bZIP-binding sites. Most notable is the partial B → A conformational transition (evidenced by the large negative Slide, positive Z_P and accompanying large-scale changes in χ, δ, and ζ values) observed in the λ complex, but not observed in the GCN4-bound structures. Furthermore, these perturbations of the crystal structure occur at G and T residues, which is not inconsistent with the nucleotides implicated by Raman spectra (Figure 7) as sites of major helical deformation in the λ repressor:operator complex in solution. These major structural distortions were not highlighted in earlier descriptions of the λ structure (*10*). The induction of non-B local structure may be associated with both specific side-chain/DNA contacts (for example, binding of Thr 2 and Lys 4 in the repressor N-terminal arm to the back of the operator) and sites adjoining a region of enlargement of the major groove required to accommodate the HTH element (for example, CTG). It is noteworthy that such distortions are not required in order for the AP-1 site to act as a template for folding and binding of the divergent α-helical arms in the bZIP complex.

On the other hand, deviations of base-pair step parameters—twist, tilt, roll, shift, slide and rise (*39*)—from values typical of B-DNA (*23*, *24*) are greater in the GCN4:AP-1 crystal complex than in the GCN4:ATF/CREB complex. This

Table 1: Raman Markers and Proposed Structural Correlations for Different Mechanisms of Protein/DNA Recognition

recognition	Raman marker (cm^{-1})[a]	proposed structural correlation
major groove	840 ↓	alteration of OPO torsion (at AT pairs)
	1376−1380 ↑	shielding of thymine $C5H_3$
	1490 → 1470	guanine N7 hydrogen bond formation
	1575 → 1580	guanine N7 hydrogen bond formation
	1720 ↓	guanine O6 hydrogen bond formation
minor groove	668 → 640	C2′-endo/anti dT to C3′-endo/anti dT
	722 → 732	C2′-endo/anti dA to C3′-endo/anti dA
	732 ↑	adenine unstacking
	785 → 796	modification of B-form DNA backbone
	819 → 850	change of α, β, γ torsions (g^-tg^+ to *ttt*)
	928 ↓	altered furanose geometry[b]
	1238 ↑	thymine unstacking
	1419 ↓	altered furanose geometry[b]
	1455 ↑	altered furanose geometry[b]
	1490 ↑	purine unstacking
	1575 ↑	purine unstacking
nonspecific[a]	732 ↑	adenine unstacking
	1490 ↑	purine unstacking
	1575 ↑	purine unstacking

[a] Arrow indicates a shift in band center (→), an intensity increase (↑), or an intensity decrease (↓) with formation of the specific protein: DNA complex. [b] Presumed diagnostic of a high degree of bending of the DNA backbone. [c] See ref *43*.

unexpected finding could simply reflect (a) the unreliability of derived parameters characterizing lower-resolution structures such as the GCN4:AP-1 complex (*38*), (b) differences between the solution vs crystal conformations of the two protein complexes, or (c) subtle influences of the individual proteins on their DNA targets. The deviations of base-pair parameters from standard values are much less pronounced in the higher resolution GCN4:ATF/CREB structure with a different half-site spacing. Past comparisons of the two crystal structures indicated that the bZIP domains were similar but that the ATF/CREB DNA site was distorted compared to the AP-1 site (*5*, *6*). Biochemical studies suggest further that such distortions, at least as monitored by electrophoretic assays of DNA bending, preexist in the DNA site and may not necessarily be imposed by the protein (*40*). The present conformational analysis corroborates the B-like character of the bound ATF/CREB site. It will be of interest to investigate by Raman spectroscopy whether such local features (as distinct from overall DNA curvature) preexist in the corresponding protein-free ATF/CREB DNA. Given that homologous bZIP domains (though not GCN4 itself) can distinguish between AP-1 and ATF DNA sites, it seems unlikely that the two sites provide—through preexisting local structure distortions—essentially equivalent templates for protein folding. It would be intriguing if among bZIP domains an induced fit is required in one complex but not the other and if the differential price in the free energy of DNA reorganization contributes to such selectivity of recognition.

Probable correlations between specific Raman difference bands (peaks and troughs in the difference signature of Figure 7) and specific geometrical parameters of B-DNA are proposed in the right-hand column of Table 1. These are presented as a working hypothesis to be further tested by ongoing Raman, UVRR, NMR, and theoretical studies of additional gene-regulatory complexes in the authors' laboratories.

CONCLUSIONS

The present results demonstrate that binding of the bZIP protein GCN4 to the AP-1 target site perturbs the vibrational Raman signature of B-DNA only minimally. In contrast, more substantial perturbations to the B-DNA Raman signature accompany binding of the phage λ cI repressor at its O_L1 target site (*10*, *16*−*18*). Further, extensive and extraordinarily large perturbations occur in the Raman signature of the B-DNA site bound by the HMG box of the hSRY transcription factor (*20*). In Figure 7, these three qualitatively distinct Raman fingerprints of protein/DNA recognition are compared with one another and with a fourth type of protein/DNA recognition, that of sequence-nonspecific binding exemplified by the single-stranded (ss) DNA-binding protein gVp of phage M13 (*43*).

The results obtained for two sequence-specific major-groove binding mechanisms—the bZIP motif in GCN4:DNA and the HTH motif in cI:DNA—indicate that sequence-specific major-groove recognition is not defined by a single Raman signature. HTH- and bZIP-binding mechanisms are each characterized by specific Raman patterns of DNA structural perturbation. Both patterns are also clearly distinguishable from the Raman difference signature of sequence-specific minor-groove recognition of hSRY−HMG:DNA. Finally, nonspecific binding of gVp to ssDNA furnishes a fourth distinctive Raman signature of protein/DNA recognition. Table 1 summarizes Raman markers associated with each binding mechanism.

The data of Figure 7 and Table 1 are of value in several respects. First, they provide an empirical database proposed to be diagnostic of different protein/DNA binding mechanisms. In principle, such a library of Raman markers could be exploited to identify the pattern of protein/DNA recognition in gene-regulatory complexes that are not amenable to high-resolution structural methods. Included among such complexes are those refractory to crystallization or containing larger nucleic-acid-binding sites. An example of this type of application is the specific complex between phage D108 Ner repressor and a 61-bp operator site, the Raman difference signature of which has been structurally characterized (*19*) on the basis of comparison with cI:DNA results. Second, the empirical database may be exploited to compare protein/DNA recognition patterns of structurally related but biologically distinct molecules, such as wild-type and mutant transcription factors (*20*) or operators (*17*). Third, these data provide a framework (Figure 8) for evaluating the sensitivity of localized vibrational modes of DNA to the specific helical parameters perturbed by protein binding. The analysis presented for the hSRY−HMG box and other minor-groove-binding proteins in the preceding paper in this issue (*20*) is an example of this type of application.

The correlations proposed in Table 1 between specific Raman difference bands (Figure 7) and specific geometrical

parameters of B-DNA (Figure 8) will be evaluated in future work.

ACKNOWLEDGMENT

We thank Dr. T. Ellenberger for the GCN4 expression plasmid and Drs. T. Chiu and R. E. Dickerson for molecular images shown in Figure 1.

REFERENCES

1. Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053−1095.
2. Kim, J. L., and Burley, S. K. (1994) *Nat. Struct. Biol. 1*, 638−653.
3. Werner, M. H., Huth, J. R., Gronenborn, A. M., and Clore, G. M. (1995) *Cell 81*, 705−714.
4. Patel, L., Abate, C., and Curran, T. (1990) *Nature 347*, 572−575.
5. Ellenberger, T. E., Brandt, C. J., Struhl, K., and Harrison, S. C. (1992) *Cell 71*, 1223−1237.
6. Konig, P., and Richmond, T. J. (1993) *J. Mol. Biol. 233*, 139−154.
7. Keller, W., Konig, P., and Richmond, T. J. (1995) *J. Mol. Biol. 254*, 657−667.
8. Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C., and Struhl, K. (1990) *Nature 347*, 575−578.
9. Kissinger, C. R., Liu, B., Martin-Blanco, E., and Kornberg, T. B. (1990) *Cell 63*, 579−590.
10. Beamer, L. J., and Pabo, C. O. (1992) *J. Mol. Biol. 227*, 177−196.
11. Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) *Nature 335*, 321−329.
12. Frederick, C. A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobson, L., Wang, B.-C., Greene, P., Boyer, H. W., and Rosenberg, J. M. (1984) *Nature 309*, 327−331.
13. Rozenberg, J., Rabinovich, D., Frolow, F., Hedge, R. S., and Shakked, Z. (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95, 15194−15199.
14. Nilges, M. (1996) *Curr. Opin. Struct. Biol. 6*, 615−623.
15. Neidle, S. (1998) *Nat. Struct. Biol. 5*, 754−756.
16. Benevides, J. M., Weiss, M. A., and Thomas, G. J., Jr.. (1991) *Biochemistry 30*, 4381−4388.
17. Benevides, J. M., Weiss, M. A., and Thomas, G. J., Jr.. (1991) *Biochemistry 30*, 5955−5963.
18. Benevides, J. M., Weiss, M. A., and Thomas, G. J., Jr. (1994) *J. Biol. Chem. 269*, 10869−10878.
19. Benevides, J. M., Kukolj, G., Autexier, C., Aubrey, K. L., DuBow, M., and Thomas, G. J., Jr.. (1994) *Biochemistry 33*, 10701−10710.
20. Benevides, J. M., Chan, G., Lu, X.-J., Olson, W. K., Weiss, M. A., and Thomas, G. J., Jr. (2000) *Biochemistry 39*, 537−547.
21. Lamba, O. P., Becka, R., and Thomas, G. J., Jr. (1990) *Biopolymers 29*, 1465−1477.
22. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R., and Schneider, B. (1992) *Biophys. J. 63*, 751−759.
23. Gorin, A. A., Zhurkin, V. B., and Olson, W. K. (1995) *J. Mol. Biol. 247*, 34−48.
24. Olson, W. K., Gorin, A. A., Lu, X.-J., Hock, L. M., and Zhurkin, V. B. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 11163−11168.
25. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) *Science 254*, 539−544.
26. Evertsz, E. M., Thomas, G. A., and Peticolas, W. L. (1991) *Biochemistry 30*, 1149−1155.
27. Chen, M. C., and Lord, R. C. (1974) *J. Am. Chem. Soc. 96*, 4750−4752.
28. Krimm, S. (1987) in *Biological Applications of Raman Spectroscopy* (Spiro, T. G., Ed.) pp 1−45, Wiley, New York.
29. Overman, S. A., and Thomas, G. J., Jr. (1998) *J. Raman Spectrosc. 29*, 23−29.
30. Overman, S. A., and Thomas, G. J., Jr. (1999) *Biochemistry 37*, 5654−5665.
31. Weiss, M. A. (1990) *Biochemistry 29*, 8020−8024.
32. Thomas, G. J., Jr., and Tsuboi, M. (1993) *Adv. Biophys. Chem. 3*, 1−70.
33. Duguid, J. G., Bloomfield, V. A., Benevides, J. M., and Thomas, G. J., Jr. (1996) *Biophys. J. 71*, 3350−3360.
34. Erfurth, S. C., and Peticolas, W. L. (1975) *Biopolymers 14* , 247−264.
35. Benevides, J. M., Stow, P. L., Ilag, L., Incardona, N. L., and Thomas, G. J., Jr. (1991) *Biochemistry 30*, 4855−4863.
36. Ansevin, A. T., and Brown, B. W. (1971) *Biochemistry 10*, 1133−1138.
37. Nishimura, Y., Tsuboi, M., Sato, T., and Akoi, K. (1986) *J. Mol. Struct. 146*, 123−153.
38. Shui, X., McFail-Isom, L., Hu, G. G., and Williams, L. D. (1998) *Biochemistry 37*, 8341−8355.
39. Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C.-S., Wang, A. H.-J., and Zhurkin, V. B. (1989) *J. Mol. Biol. 208*, 787−791.
40. Hockings, S. C., Kahn, J. D., and Crothers, D. M. (1998) *Proc. Natl. Acad. Sci. 95*, 1410−1415.
41. Dickerson, R. E., and Chiu, T. K. (1998) *Biopolymers 44*, 361−404.
42. Thomas, G. J., Jr., and Agard, D. A. (1984) *Biophys. J. 46*, 763−76.
43. Benevides, J. M., Terwilliger, T. C., Vohník, S., and Thomas, G. J., Jr. (1996) *Biochemistry 35*, 9603−9609.
44. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, Jr., E. E., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.

BI990053X