P22 c2 Repressor−Operator Complex: Mechanisms of Direct and Indirect Readout‡

Derrick Watkins,§ Chiaolong Hsiao,§ Kristen Kruger Woods,§ Gerald B. Koudelka,*,|| and Loren Dean Williams*,§

School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332-0400, and Department of Biological Sciences, 607 Cooke Hall, State University of New York at Buffalo, Buffalo, New York 14260

Received September 6, 2007; Revised Manuscript Received November 13, 2007

ABSTRACT: The P22 c2 repressor protein (P22R) binds to DNA sequence-specifically and helps to direct the temperate lambdoid bacteriophage P22 to the lysogenic developmental pathway. We describe the 1.6 Å X-ray structure of the N-terminal domain (NTD) of P22R in a complex with a DNA fragment containing the synthetic operator sequence $[d(ATTTAAGATATCTTAAAT)]_2$. This operator has an A-T base pair at position 9L and a T-A base pair at position 9R and is termed DNA^{9T}. Direct readout: nondirectional van der Waals interactions between protein and DNA appear to confer sequence-specificity. The structure of the P22R NTD−DNA^{9T} complex suggests that sequence-specificity arises substantially from lock-and-key interaction of a valine with a complementary binding cleft on the major groove surface of DNA^{9T}. The cleft is formed by four methyl groups on sequential base pairs of 5′-TTAA-3′. The valine cleft is intrinsic to the DNA sequence and does not arise from protein-induced DNA conformational changes. Protein−DNA hydrogen bonding plays a secondary role in specificity. Indirect readout: it is known that the noncontacted bases in the center of the complex are important determinants of affinity. The protein induces a transition of the noncontacted region from B-DNA to B′-DNA. The B′ state is characterized by a narrow minor groove and a zigzag spine of hydration. The free energy of the transition from B- to B′-DNA is known to depend on the sequence. Thus, the observed DNA conformation and hydration allows for the formulation of a predictive model of the indirect readout phenomenon.

Recognition of duplex DNA[1] sequences by proteins plays critical roles in mediating transcription, recombination, replication, and DNA packaging. DNA recognition is accomplished by a combination of two broadly defined mechanisms: direct readout and indirect readout.

Direct readout of a DNA sequence is the sensing of base pair identity by direct hydrogen bonding and van der Waals interactions with functional groups on the DNA base pairs (*1*, *2*). Here, we describe a novel mechanism of direct readout that utilizes shape complementarity between the major groove of the DNA and the protein reading head. A sequence-specific cleft, formed by the methyl groups of thymines on the major groove surface, matches the shape of the recognition helix. Simple nondirectional van der Waals interactions between protein and DNA appear to confer sequence-specificity.

Indirect readout of DNA sequences is the sensing of base pair identity without direct base−protein contact. Indirect readout utilizes DNA sequence-dependent deformability (i.e., see ref *3*). Consequently, indirect readout allows the sensing of the DNA sequence at a distance.

434 repressor (*4*−*6*), TATA binding protein (TBP, refs *7* and *8*−*10*), and trp repressor (*11*, *12*) are among the first proteins shown to indirectly read (recognize) portions of their DNA binding sites. DNA binding proteins that have been known to use indirect readout for recognition now extend over the entire tree of life and include met repressor (*13*), IHF/Hbb (*14*, *15*), c-myb (*16*), mar A (*17*), papillomavirus E2 protein (*18*, *19*), estrogen receptors (*20*), CAP (*21*), HincII restriction endonuclease (*22*), P22R (*6*, *23*), and Ndt80 (*24*).

Indirect readout exploits the complex relationships between DNA sequence, conformation, deformation, hydration, and electrostatics. Despite the prevalence and biological importance of indirect readout, predictive mechanistic models of this phenomenon are not available. Here, we describe a DNA−protein structure that reveals the mechanism of indirect readout and yields a general predictive model. The hydration and conformation of portions of this structure are consistent with a protein-induced transition from B-DNA to B′-DNA. The free energy of this change in the DNA state is known to vary with sequence (*25*). The structure provides a molecular level explanation for indirect readout: a protein-induced alteration in the noncontacted region of the DNA that is supported by interactions between minor groove functional groups of DNA and water molecules and presumably cations.

To direct the temperate lambdoid bacteriophage P22 to the lysogenic developmental pathway, P22R must recognize and bind six naturally occurring binding sites with different affinities (*26*, *27*). The six sites are arranged within the phage

‡ Atomic coordinates have been deposited in the Protein Data Bank (PDB ID 2R1J and RCSB ID RCSB044309).

* Corresponding authors. (G.B.K.) E-mail: koudelka@acsu.buffalo.edu; phone: (716) 645-2363, ext. 158; fax: (716) 645-2871. (L.D.W.) E-mail: loren.williams@chemistry.gatech.edu; phone: (404) 894-9752; fax: (404) 894-7452.

§ Georgia Institute of Technology.

|| State University of New York at Buffalo.

[1] Abbreviations: CAP, catabolite activating protein; DNA, deoxyribonucleic acid; DNA^{9T}, $[d(ATTTAAGATATCTTAAAT)]_2$; NTD, N-terminal domain; TBP, TATA box binding protein; trp, tryptophan; P22R, bacteriophage c2 repressor; Rel, relative.

10.1021/bi701826f CCC: $40.75
Published on Web 02/01/2008

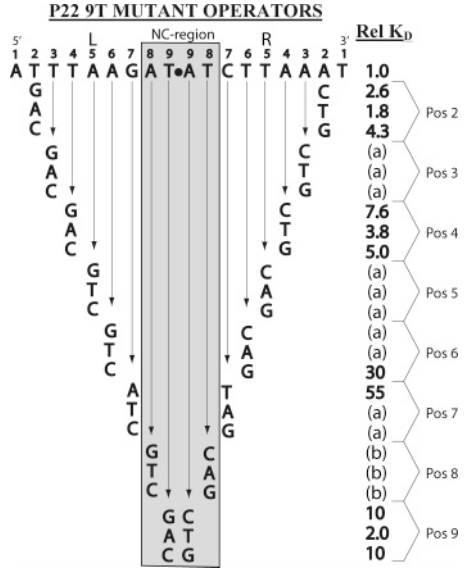

FIGURE 1: Consensus P22R binding sequence (synthetic) and mutants of the synthetic binding site. Binding affinities of P22R are expressed relative to that of the consensus synthetic operator. The Rel K_D is the relative dissociation constant at 100 mM KCl normalized to the tightest binding synthetic operator, 9T. The apparent K_d (K_D) is defined as the concentration of repressor monomers needed to half maximally occupy a binding site. Under the conditions of these experiments, the P22 repressor would be anticipated to be 100% monomeric. Hence, any differences in apparent K_D reflect changes in the stability of the repressor−DNA complexes. The value used for normalization is $1 = 1.6 \times 10^{-8}$ M. The gray bar highlights the positions of the noncontacted base pairs. "a" indicates that the repressor binds to the operator only nonspecifically. Nonspecific binding is observed at concentrations $<2 \times 10^{-6}$ M. "b" indicates that data are not available. The sequences of natural P22 binding sites are given in ref *30*.

into two operator regions, O_R and O_L (*28*, *29*). The P22R binding sites are designated O_R1, O_R2, and O_R3 in the O_R operator and O_L1, O_L2, and O_L3 in the O_L operator. The sequences of the naturally occurring sites (*30*) are partially symmetric. The sequences of the symmetrically arrayed outermost base pairs in these sites are highly conserved, but the sequences of the innermost bases are divergent (*23*, *29*).

Biochemical experiments have suggested that conserved bases in the naturally occurring sites (*30*) are contacted by P22R (*30*, *31*). Similar experiments indicate that the variable central bases are not contacted by P22R. Yet, changes in the sequence of these noncontacted bases alter the operator's affinity for protein (*23*, *30*−*32*). For example, T-A → C-G transitions of operator 9T at positions 9L and 9R (Figure 1) decrease the P22R affinity by 10-fold (*23*). Hence, the P22R protein recognizes its specific DNA sites by directly contacting highly conserved bases in the outer regions of the sites and discriminates between sites by indirect readout, by sensing the central noncontacted sequence.

To decipher mechanisms of both direct and indirect readout, we solved the X-ray structure of the P22R N-terminal DNA binding domain (P22R NTD, residues 1−68) in complex with the synthetic 9T operator (DNA^{9T}). DNA^{9T} has a T-A base pair at position 9L and an A-T base pair at position 9R (see Figure 1). The structure of the complex allows us to extend and refine the binding model of Koudelka et al. (*31*). This model is based on extensive mutagenesis and biochemical data (Figure 1), the NMR structure of P22R alone (*33*), and homology modeling, using the 3-D structures of 434 repressor−operator complexes (*34*−*36*) as templates. The model predicts specific contacts and attempts to account for highly conserved and variable positions of the operator. The 3-D structure described here allows us to test and extend Koudelka's model.

MATERIALS AND METHODS

Construction of Plasmid Expressing P22R NTD. A DNA fragment encoding the NTD fragment of P22R was created using PCR to insert a stop codon at position 69 of the intact P22R gene. We amplified a 450 bp fragment from pTP125 (*37*) using primers obtained from the CAMBI Nucleic Acid Facility (University at Buffalo). The sequences of the primers were 5′-GTTTTTTGCGTCGACATCATA-3′ and 5′-CGTTTGTCTGACGTCAATCTCCTTTC-3′. This fragment, containing the *tac* promoter (*38*) fused to the P22R gene fragment, was cleaved with *Aat* II and *Sal* I, gel purified, ligated in pUC18 cut with the same enzymes, and transformed into *Escherichia coli* strain XA90 (*39*). The resulting plasmid expresses a protein comprised of the first 68 amino acids of P22R.

Purification of P22R NTD. *E. coli* XA90 cells (*39*) bearing a P22R NTD expression plasmid were grown with aeration at 37 °C in 3 L of Luria broth supplemented with 0.1 mg/mL ampicillin to an A_{600} of approximately 0.6. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to 1 mM, and growth at 37 °C was continued for 4 h. The cells were harvested by centrifugation and resuspended in 20 mL of lysis buffer (200 mM NaCl and 50 mM EDTA; 100 mM Tris, pH 7.4) with 0.5 mL of 20 mg/mL phenyl methyl sulfonyl fluoride (PMSF). The cells were lysed in a French pressure cell (18 000 psi) at 4 °C. The lysate was diluted with 100 mL of cold lysis buffer. Cellular debris was precipitated by centrifugation at 4000*g*. Nucleic acids were precipitated by dropwise addition of polyethyleneimine to a final concentration of 0.6% at 4 °C and mixed for 15 min, followed by centrifugation at 10 000*g*. Soluble proteins were precipitated with 40% (w/v) ammonium sulfate and dialyzed at 4 °C against three 1 L changes of standard buffer (SPB-50 mM $NaPO_4$, pH 6.8, 10 mM EDTA) supplemented with 50 mM NaCl. The dialyzate was loaded onto three Econo-Pac carboxymethyl (CM) cartridges (Bio-Rad) and eluted with a linear salt gradient from 0.1 to 1 M NaCl in SPB over 70 min. Fractions containing the repressor were pooled and precipitated with 40% ammonium sulfate. Following precipitation, the protein-containing pellet was dissolved in SPB and 100 mM NaCl and loaded onto a 1 cm × 50 cm Sephacryl S100 (Amersham Biosciences) size exclusion column. The column was developed with the same buffer at a flow rate of 0.3 mL/min. Fractions containing P22R NTD were pooled, and the protein was precipitated with 40% ammonium sulfate. The dissolved pellet was dialyzed against SPB supplemented with 150 mM NaCl and 20% glycerol. The protein was flash frozen in liquid nitrogen and stored at −70 °C.

DNA Purification. DNA oligonucleotides were obtained from Integrated DNA Technologies. The oligonucleotides were purified with denaturing polyacrylamide gel containing 7 M urea and TBE (89 mM Tris borate, pH 8.4, 1 mM EDTA) and subsequently dissolved in TEN (10 mM Tris,

pH 8.0, 50 mM NaCl, 1 mM EDTA). To form the double-stranded protein binding site, equimolar amounts of the complementary strands were mixed, heated to 85 °C for 60 s, and slow-cooled over 4 h to anneal. Double-stranded DNA was separated from the single strands by electrophoresis on 12% polyacrylamide gels in 1X TBE.

Cocrystallization of P22R NTD with DNA^{9T}. Crystals were grown at 4 °C by the hanging drop vapor diffusion method. The initial crystallization solution contained 0.42 mM P22R NTD, 0.42 mM duplex d(5′-TATTTAAGATATCTTAAATG-3′)−d(5′-CATTTAAGATATCTTAAATA-3′), 45 mM Tris·HCl (pH 7.8), 19 mM NaCl, 1.9 mM glycerol, 11% PEG 400, 4.5 mM LiCl, 2.3 mM $MgCl_2$, and 0.91% MPD in a volume of 5.3 μL. The crystallization solution was equilibrated against a reservoir of 100 mM Tris·HCl (pH 7.8), 25% PEG 400, 10 mM LiCl, 5 mM $MgCl_2$, and 2% MPD. Crystals grew to a size of approximately 0.2 mm × 0.2 mm × 0.2 mm within 1 day. A single crystal was collected from the crystallization solution in a nylon loop and was flash frozen in liquid nitrogen.

Data Collection. Diffraction data were collected on SER-CAT beamline 22 ID at the Advanced Photon Source, Argonne National Laboratory with a MAR300 CCD detector using radiation 0.99997 Å in wavelength and a crystal to detector distance of 170 mm. Images were collected with a 1 s exposure time and a 1° oscillation. The crystal was maintained at 110 °K. A total of 347 031 reflections were indexed and integrated with DENZO (*40*) and reduced to 55 132 unique reflections in space group $P4_3$ in the resolution range of 35−1.53 Å. Data were scaled with the program SCALEPACK (*41*).

Refinement. Initial phases were determined by molecular replacement with the program CNS (*42*) using data from 15 to 3 Å and the 434 repressor−O_R2 complex (PDB ID 1RPE) as a search model. All nonidentical residues (P22R vs 434 repressor) in the search model were mutated to alanine. The 434 repressor NTD shares ∼30% sequence identity with the P22R NTD. A rigid body refinement in CNS with numerous cycles of model building based on $|2F_o - F_c|$ and $|F_o - F_c|$ maps was performed in the program O (*43*). The next step was refinement in CNS using the *R*-factor and R_{free} as guides. To reduce model bias, simulated annealing was performed during each stage of refinement with a starting temperature of 5000 °C and 25 °C steps in the early rounds of refinement and a 2000 °C starting temperature and 50 °C steps in later stages. Water molecules were added iteratively to the model, followed by cycles of refinement. Composite omit $|2F_o - F_c|$ maps were generated with CNS, omitting 5% of the model in each round of refinement, with an initial annealing temperature of 2500 K. The final refined model of the P22R NTD (residues 1−68) in a complex with DNA^{9T} was refined to 1.53 Å resolution, with an *R*-factor of 20.35% and a R_{free} value of 22.53%. Data collection and refinement statistics are given in Table 1.

Electron Density and Model Quality. The backbone atoms of the protein, as well as the side chain atoms, are well-ordered and clearly observable in the $|2F_o - F_c|$ Fourier maps with only a few exceptions occurring for long surface side chains and terminal residues. The N-terminal residues (1 and 2) are apparently disordered, and their density is not observed. These four residues (two per monomer) are omitted from the model. Ill-defined density was observed for residue 3 and C-terminal residue 68. These residues are built in at half occupancy, which gives the best fit to the data, using the *R*-factor and R_{free} values, as well as the sum and difference density maps as guides.

Table 1: Data Collection and Refinement Statistics

X-ray source	SER-CAT beamline 22-ID APS
wavelength (Å)	0.99997
detector	MAR CCD 300
resolution range (Å)	35.00−1.53
mosaic spread (deg)	∼0.7
space group	$P4_3$
unit cell params (Å)	
a	64.059
b	64.059
c	101.604
total reflns	347031
unique reflns	55132
multiplicity	5.8 (4.7)[a]
completeness (%)	99.4 (99.3)
highest resolution shell (Å)	1.57−1.53
$I/\sigma(I)$	47.1 (2.12)
R_{sym} (%)	6.9 (65.9)
reflns (working/test set)	53765/6068
no. of non-H protein atoms	2 × 514
no. of non-H DNA atoms	814
no. of water atoms	330
$R_{crystal}$	20.35%
R_{free}	22.53%
av *B*-factor	
protein	31.55 (L), 25.69 (R)
DNA	29.43
water	42.62
rmsd from ideal	
bond length (Å)	0.0099
bond angle (deg)	1.33

[a] Number in parentheses indicates the value for the highest resolution shell.

The electron density of DNA is sharp and well-defined throughout with exceptions at the terminal base pairs. The R half site has better-defined density, lower thermal factors, and more observable solvent molecules than the L half site. The conformation of the DNA oligomer is not symmetric in that one of the terminal base pairs is a G·C base pair and the other is an A·T base pair. The *R*-factor and R_{free} values as well as the difference density indicate a best fit with the G·C base pair in the R half site terminus and the A·T base pair in the L half site terminus. The terminal base pairs are probably characterized by multiple conformations and mixed base pair occupancy. The electron density is not well-defined at either terminus. The thermal factors of the L half site terminus base pair are highest. Attempts to model multiple conformations were not justified by fitting statistics. Therefore, only single, apparently predominant occupied position and base IDs were modeled for the terminal base pairs.

A Ramanchandran plot (*44*) calculated with Procheck (*45*) showed that 94% of residues fall in the most favored regions of the backbone conformation, while 6% of residues fall in allowed regions. No residues fall in the generously allowed or forbidden region of the Ramanchandran plot.

RESULTS

Global Complex. The electron density maps of the complex are clean, continuous, and detailed (Figure 2). P22R NTD binds to DNA^{9T} as a dimer (Figure 3). The two subunits

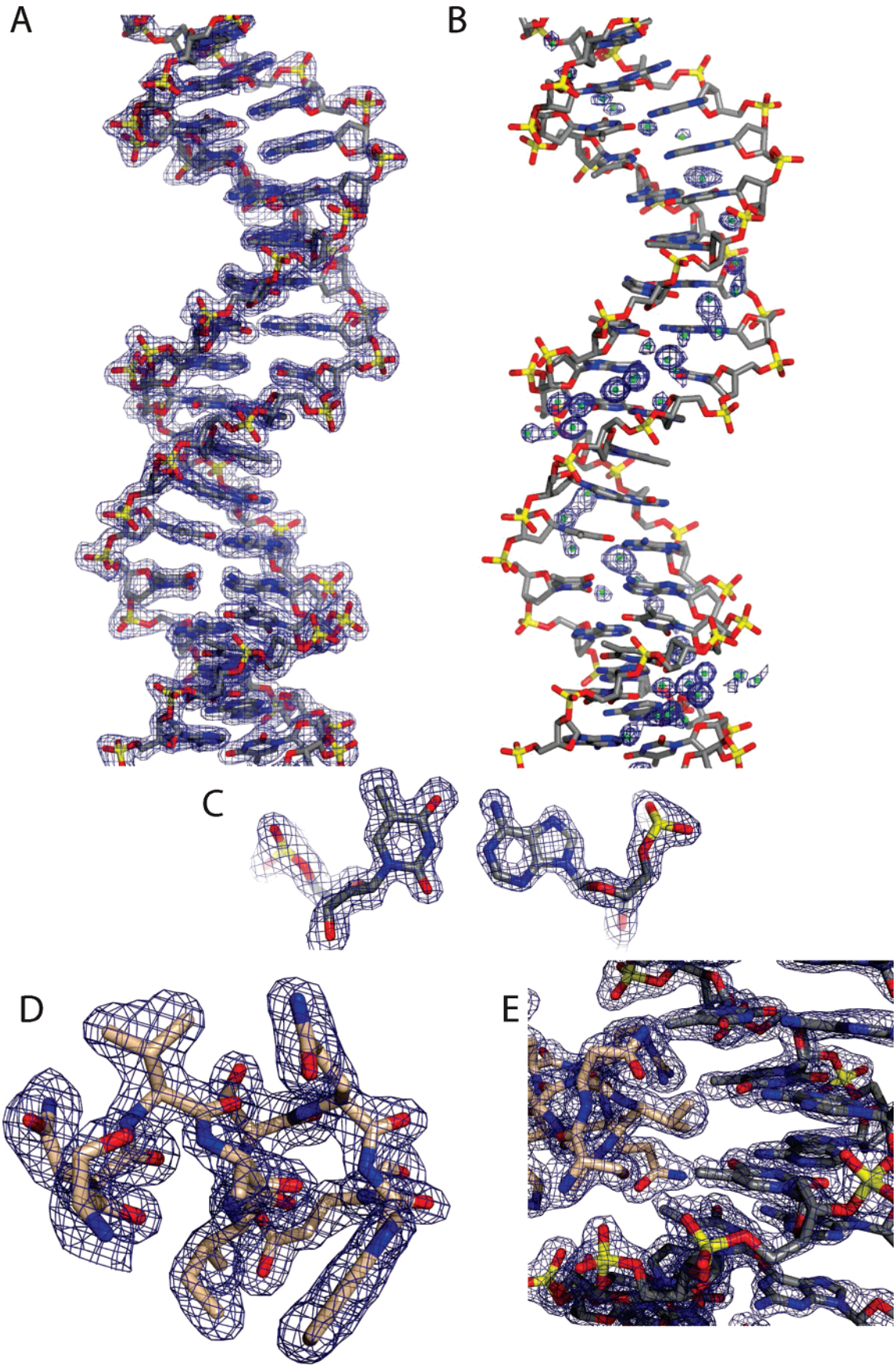

FIGURE 2: Electron density defining that the DNA, protein, and solvent are of high quality. Shown here is (A) electron density surrounding the DNA duplex contoured at 1.5 σ, (B) electron density surrounding the minor groove solvent contoured at 1.3 σ, (C) electron density surrounding the base pair T(9)-A(32) contoured at 1.3 σ, (D) electron density surrounding the amino acids of the recognition helix (residues 31−40) contoured at 1.5 σ, and (E) electron density surrounding the DNA and R recognition helix contoured at 1.5 σ. The maps are annealed/refined $|2F_o - F_c|$ composite omit maps.

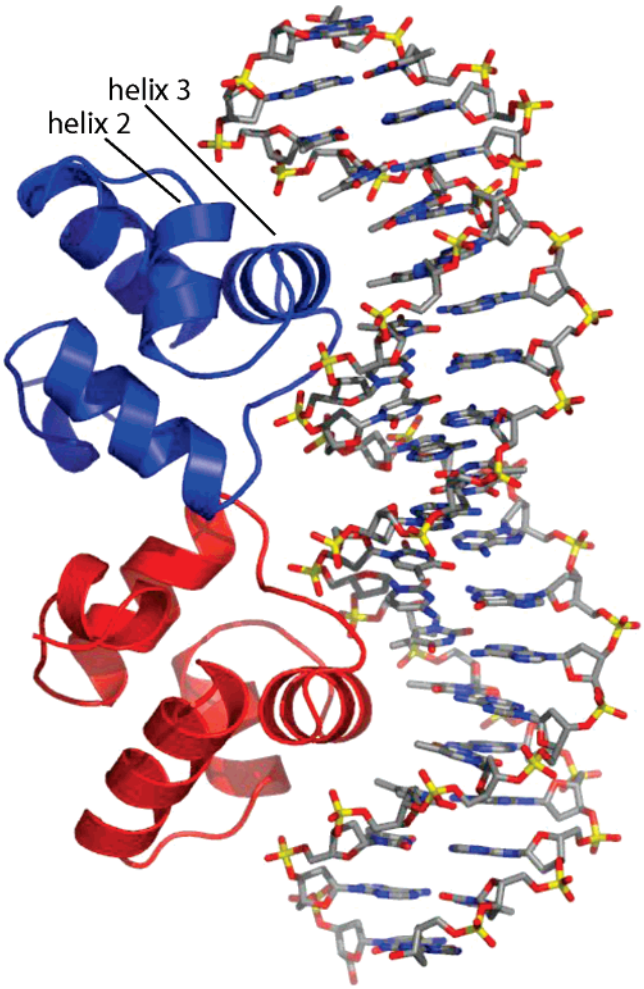

FIGURE 3: Global view of the DNA−protein complex. The R subunit is represented by the red ribbon. The L subunit is represented by the blue ribbon. DNA is represented as sticks with C, gray; N, blue; O, red; and P, yellow.

of the dimer are related by an approximate 2-fold rotational axis that is roughly perpendicular to the DNA helical axis, bisecting the 18 base pair recognition sequence. Protein interactions with the two DNA half sites are symmetric about pseudo-2-fold in the center of the operator. With a few exceptions, this symmetry extends down to the level of solvent locations and interactions. Hence, the description here is given for the structure and interactions of the R half site (Figure 4) but applies equally well to the L half site. The observed symmetry argues against the influence of specific lattice forces on conformation or interactions.

Each protein monomer contains five α-helices made up of residues 6−17 (helix 1), 21−28 (helix 2), 32−40 (helix 3), 47−57 (helix 4), and 61−65 (helix 5). The X-ray structure of P22R NTD in the DNA complex is consistent with the Wüthrich NMR structure of the protein alone (*33*). The protein forms a globular structure with hydrophobic residues in the core and charged and polar residues on the exterior, exposed to solvent and to DNA. As anticipated (*31*, *33*, *46*), P22R NTD contains a helix-turn-helix (HTH) motif made up of helices 2 and 3. The N-terminus of helix 3, the DNA recognition helix, is further from the central noncontacted region than the C-terminus. Helix 3 is the most ordered region of the polypeptide as indicated by thermal factors and electron density maps. The protein is less ordered in the helices of the N and C termini, with solvent-exposed residues showing the most disorder.

Global DNA Conformation. Perturbations in base pair roll and propeller twist induce complementarity of the global shape of the DNA with the shape of the protein. An overall DNA curvature of 16° allows correct alignment of successive major groove faces with recognition helices (Figure 3). The overall conformation of DNA^{9T} is the B-form (Tables 3S−5S), with directions and positions of axial deviations from linearity allowing the DNA to successfully transit from one recognition helix to the next. Deviations from linearity of the axis are associated with positive rolling between base pairs 5 and 6 (9°) and between base pairs 6 and 7 (11°).

Noncontacted Region: Induced B′ State. The noncontacted region of DNA^{9T} in the P22R NDT complex is in the B′ state as indicated by a spine of hydration and an especially narrow minor groove (Figure 5). The B′ state is not normally observed in non-A-tract DNA sequences (ATAT is not an A-tract). However, induction of the B′ state in ATAT sequences can be caused by crystal lattice forces (*47*). It appears that the B′ state of the noncontacted ATAT sequence of DNA^{9T} is caused here by protein binding. For comparison, Figure 6 shows a superimposition of the ATAT elements in the unperturbed B conformation, a lattice-induced B′ state, and our protein-induced B′ state. The minor groove is narrow in both induced B′ state fragments and is wider in the unperturbed state.

Continuous spines of hydration are observed in both the contacted A-tract region of DNA^{9T} comprised of positions R1 through R4 and in the noncontacted region (positions 8L-9L and 8R-9R), which does not contain an A-tract sequence. The base of the spine is a series of water molecules that forms bridges between hydrogen bond acceptors of DNA bases (Figure 5A,B). A secondary layer forms bridges between adjacent waters in the primary layer. At the central TpA step, water 5 bridges two N3 atoms [the N3 of A(11) and the N3 of A(31)]. This N3−W5−N3 bridge at the TpA step is not a component of A-tract hydration (TpA steps are not allowed in A-tracts). However, the N3−W5−N3 interaction very closely mimics the bridging interactions of an A-tract. Like water molecules at the ApT steps of A-tracts (*48*), W5 is six-coordinate, interacting with two cross-strand hydrogen bond acceptors on the floor of the groove, two O4′ atoms of the deoxyribose backbone, and two water molecules in the secondary layer of the spine (Figure 5B).

The secondary hydration layer of the noncontacted region minor groove is orderly and well-formed and supports the characteristic third layer of hydration. The hydration of the L end of the operator is similar but appears to be less extensive than that of the R end. This difference is likely due to the higher overall disorder of the L half site, which is indicated in the electron density and the higher overall thermal factors.

We assume that there are counterions within the noncontacted minor groove. However, the differentiation of cations such as Na^+ from water molecules by X-ray diffraction is problematic because of mixed and partial occupancy, similar scattering properties of Na^+ with water, and positional disorder. All solvent peaks are assigned as water molecules, although it is likely that some positions are at least partially

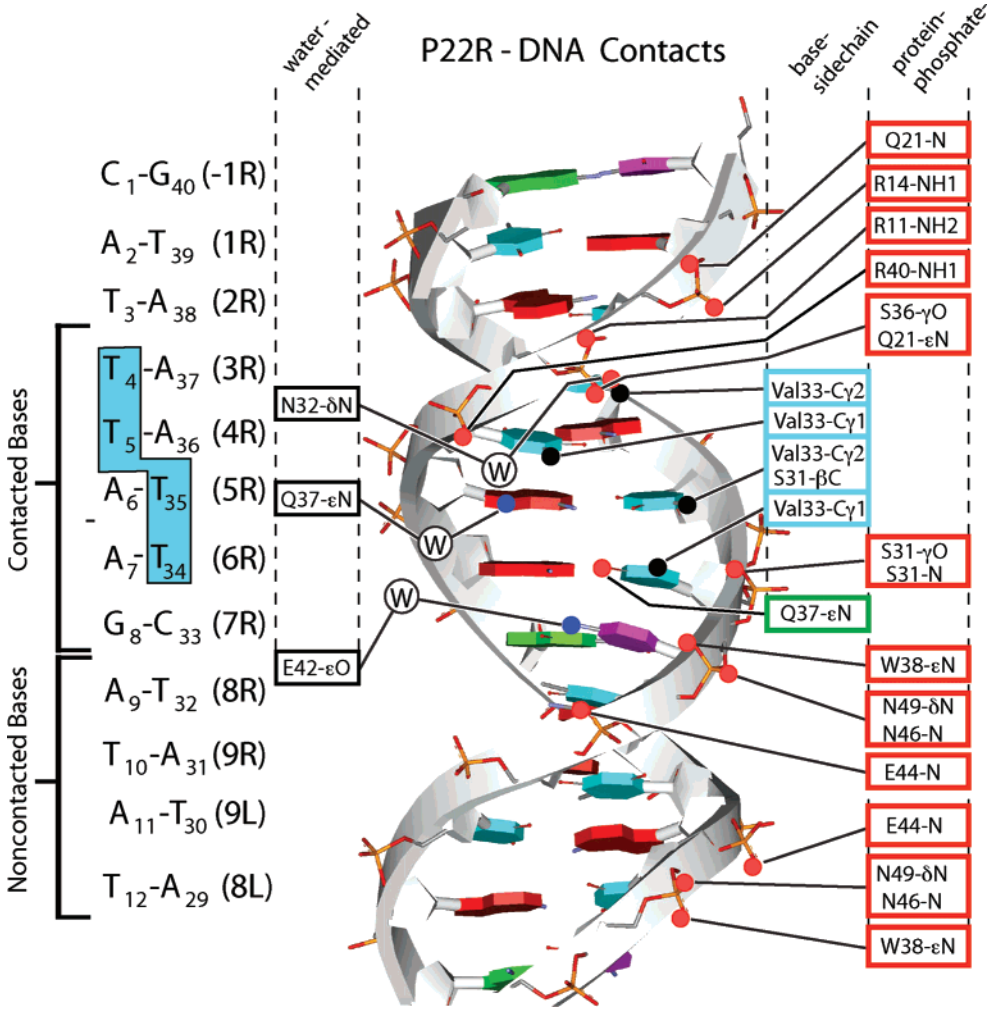

FIGURE 4: Schematic diagram of P22R NTD−DNA^{9T} interactions. Operator positions are numbered (1R, 2R, etc.) and labeled as in the PDB file (subscripts). The thymine residues that form the valine cleft are highlighted in cyan. Protein atoms that interact with DNA^{9T} are in rectangles. The left-hand column gives water-mediated contacts, the central column gives base-specific contacts, and the right-hand column gives protein−phosphate contacts. The DNA bases T (cyan), A (red), G (green), and C (purple) are shown in cartoon form. The backbone is highlighted by a gray band.

occupied by cations (*49*−*51*). Involvement of cations in the indirect readout of noncontacted bases of P22R is indicated by biochemical studies (*23*, *32*).

Narrow Minor Groove in the Noncontacted Region. The global minimum in the minor groove width is at the center of the noncontacted region (positions L9 and R9, Figure 5C). The minor groove is wide from R8 through R5 and then narrows from R4 through R1 (the A-tract).

Tunnel. The DNA^{9T} minor groove and the repressor combine to form a tunnel (Figure 7). The floor of the tunnel is the minor groove edges of the bases of residues 11−13 and 31−32. The walls are the sugar−phosphate backbone. The ceiling of the tunnel is the dimer interface (i.e., the loop region connecting helices 3 and 4 and the N-terminal region of helix 4). The tunnel is filled with solvent molecules. The floors, walls, and ceiling appear to be impermeable to water and ions. Hence, solvent molecules can enter and exit the tunnel only by traversing along the minor groove.

DNA Base−Protein Contacts. The protein side chain−DNA base interactions (Figure 4) involve helix 3 as well as several residues on either side of helix 3. These base−side chain interactions involve the outer seven base pairs of the operator half site. The two recognition helices, one from each monomer, contact consecutive major grooves, on a common helical face. The protein side chains of residues 31−33, 37, and 42 contact the functional groups of bases in the contacted region of the operator. Each of these protein residues contacts a base of the operator either directly, as with 31, 33, and 37, or via a water-mediated contact as with 32 and 42. Contacts made by these amino acids, which appear to be base-specific, are supported by buttressing interactions with other residues in the recognition helix and by Ser 31, directly preceding the recognition helix. A complex network of supporting and solvent-mediated interactions helps to explain the pleiotropic effects that changing residues in the recognition helix has on DNA binding specificity (*30*−*32*).

Valine Cleft. Valine 33, the second residue of the recognition helix, is inserted into a prominent cleft, which we call the valine cleft, in the major groove of the DNA (Figure 8). The shape of the cleft is complementary to that of the valine side chain. Interactions of Val33 within this pocket appear to be important for positioning the protein relative to the DNA and may confer thermodynamic stability and specificity.

The valine cleft is composed of two opposing sets of major groove methyl groups contributed by A-T base pairs of the

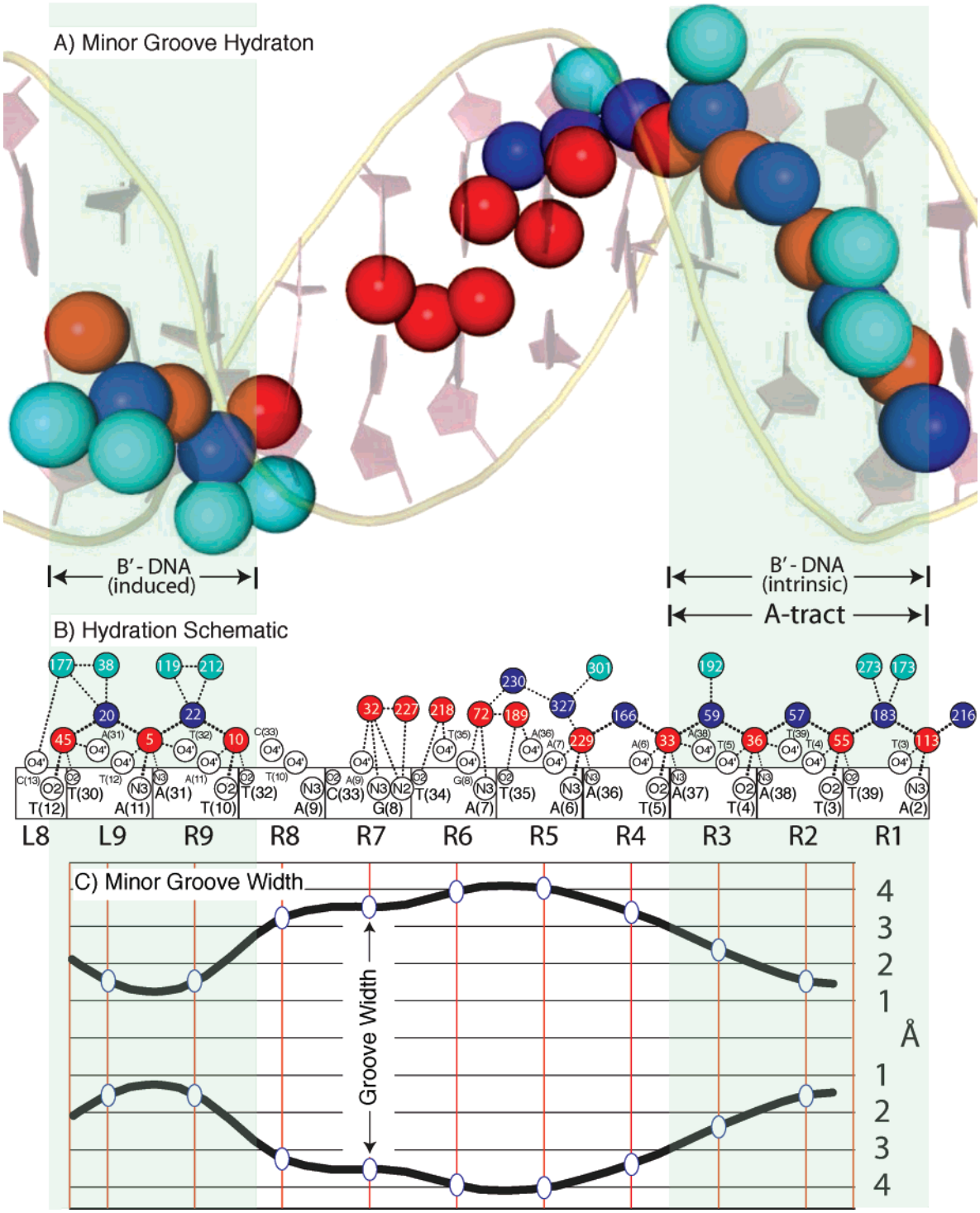

FIGURE 5: B′ state: induced and intrinsic. The R half site of the P22R NTD−DNA^{9T} complex, panels A−C are aligned, with a green background indicating the regions of induced and intrinsic B′-DNA. (A) Minor groove hydration. The primary hydration layer, composed of water molecules in contact with DNA bases, is red. The secondary hydration layer, composed of water molecules in contact with the primary layer, is blue. The tertiary hydration layer, composed of water molecules in contact with the secondary layer, is cyan. (B) Schematic representation of the minor groove hydration, showing the zigzag spine of hydration in the induced (noncontacted region) and intrinsic B′ regions. The color scheme is the same as in panel A. Hydrogen bonds are indicated by dashed lines. (C) Minor groove width profile, showing the narrow minor groove of the induced and intrinsic B′ regions. Each line represents one strand, so that the distance from line to line is the width of the minor groove.

sequence 5′-TTAA-3′. The 5-methyl groups of positions 3 and 4 form one hydrophobic wall with contributions from the methyl group of the T-A at position 2. The 5-methyl groups of positions 5 and 6 form a second hydrophobic wall. The cleft appears to be a natural consequence of the TTAA sequence element and is not induced by protein binding. An empty cleft with similar dimensions is observable in an unbound DNA fragment with the same 5′-TTAA-3′ sequence element (NDB entry BD0051 (*47*)). In DNA^{9T}, negative propeller twists (−11 to −16°) of positions 2−7 (except position 5 with a propeller twist of −3°) facilitate collapse of methyl groups on opposing sides of the major groove

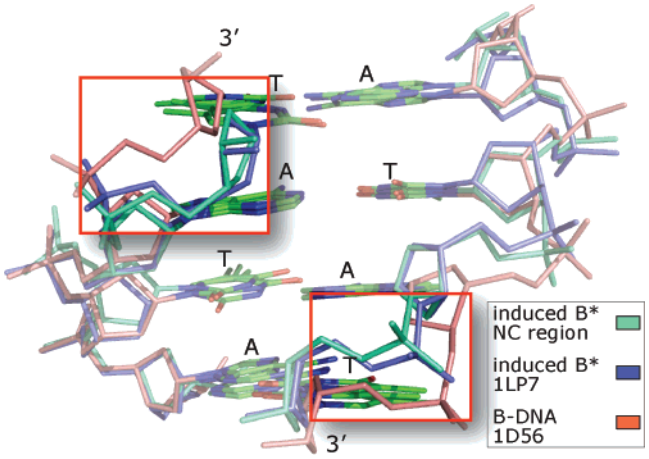

FIGURE 6: Superimposition of ATAT segments in two DNA fragments in which the B′ states have been induced, along with DNA, in the unperturbed B state. Protein-induced B′ state (P22R NTD−DNA^{9T}) is cyan, lattice-induced B′ state is blue (PDB ID 1LP7), and unperturbed B state is violet (PDB ID 1D56). The red box highlights the narrow minor groove of the two B′-DNA fragments, in comparison to the unperturbed B-DNA.

toward a common center, decreasing their relative offset along the helical axis. Similar propeller twisting is observed in the free cleft (BD0051). The natural propeller twist and base pair roll associated with the 5′-TTAA-3′ sequence are important in forming the cleft in both the free and the bound DNA fragments. The floor of the cleft is hydrophilic and is formed by the O2(T) and N6(A) atoms of base pairs 4 and 5 (Table 1S). While desolvation of the hydrophilic floor of the cleft is expected to be thermodynamically unfavorable, the desolvation of the hydrophobic Val33 side chain and the walls of the cleft would be favorable.

On one end of the cleft, Val33 is buttressed by Asn32 and Ser31, near the N-terminus of the recognition helix. The Oγ of Ser31 donates a hydrogen bond to the O1P of thymine at position 6 (2.7 Å) and accepts a hydrogen bond from the NH of Ala34 (3.1 Å). The Cβ and Oγ of Ser31 form van der Waals contacts with the 5-methyl of thymine at position 5 (3.7 Å). This interaction network appears to help stabilize the position of the protein relative to the DNA, helping to fix Val33 within the cleft.

On the other end of the cleft, closer to the center of the operator, Val33 is buttressed by Gln37 and Ser36. In the single direct hydrogen bond (per half site) between the protein and the DNA base, the ϵ-NH_2 of Gln37 forms a hydrogen bond with the O4 of thymine at position 6 (3.1 Å). Several water-mediated hydrogen bonds link ϵ-NH_2 of Gln37 to the DNA, as described next. The side chain of Gln37 has a well-defined density and low thermal factors (~23 Å^2 for Gln37 of the L chain vs ~32 Å^2 average for the L chain and ~17 Å^2 for Gln37 of the R chain vs ~26 Å^2 average for the R chain), suggesting that the water-mediated interactions restrain the position of the side chain. The position and rotameric form of Gln37 are such that the side chain methylene group contacts the Cγ of Val33 (3.9 Å). Similarly, the methylene group (Cβ) of Ser36 also contacts the Val33 side chain (4.2 Å). The Oγ of Ser36 forms two hydrogen bonds: one to the Nϵ (3.0 Å) and one to the NH2 (3.1 Å) of Arg40. The NH1 of Arg40 in turn forms a hydrogen bond with the Oϵ2 of Glu42 (3.2 Å).

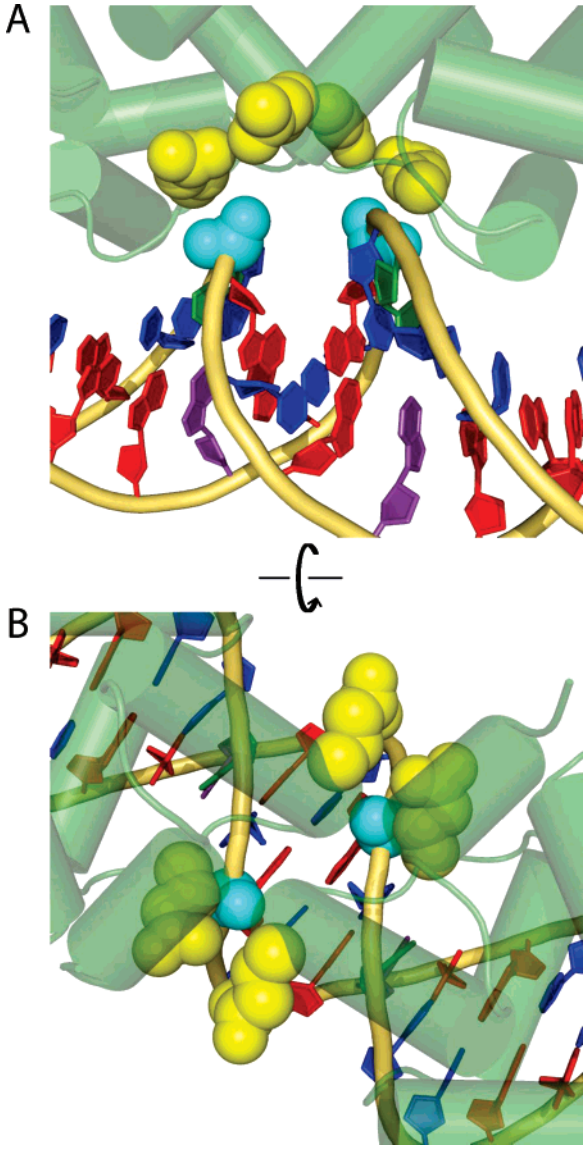

FIGURE 7: Minor groove tunnel. The tunnel is made up of the minor groove of the DNA and dimer interface of the protein. (A) View down the tunnel, along the minor groove. (B) View through the protein, across the minor groove. The protein is green with the cylinders representing the α-helices. The DNA is red (A), blue (T), purple (G), and green (C). The sugar−phosphate backbone is a brown tube. Glutamates 44 and 48 (yellow) and the proximal phosphate groups (cyan) are shown in space filling representation, emphasizing the high electronegative potential of the tunnel.

Protein−Phosphate Contacts. Each monomer of the P22R NTD forms three salt bridges (hydrogen bonds involving cationic side chains and phosphate oxygens) and nine additional hydrogen bonds with phosphate oxygens of DNA^{9T}, giving 12 direct polar contacts with phosphate oxygens (Figure 4). With one exception, all protein to phosphate contacts correspond to those identified by phosphate ethylation interference (*52*). Salt bridges from NH2 of Arg11 to the O2P at position 3 [2.9 Å, T(4)] and from NH1 of Arg14 to O2P at position 2 [3.2 Å, T(3)] help anchor helix 1 to DNA^{9T}. The recognition helix engages in four phosphate contacts. The Oγ of Ser36 contacts an O1P at position 3 [2.6 Å, T(4)], the Nϵ of Trp38 contacts an O1P at position 7 [2.8 Å, C(33)], the NH1 of Arg40 contacts an O1P at position 4 [3.4 Å, T(5)], and Ser31 makes contacts as described previously. The only phosphate contacts from

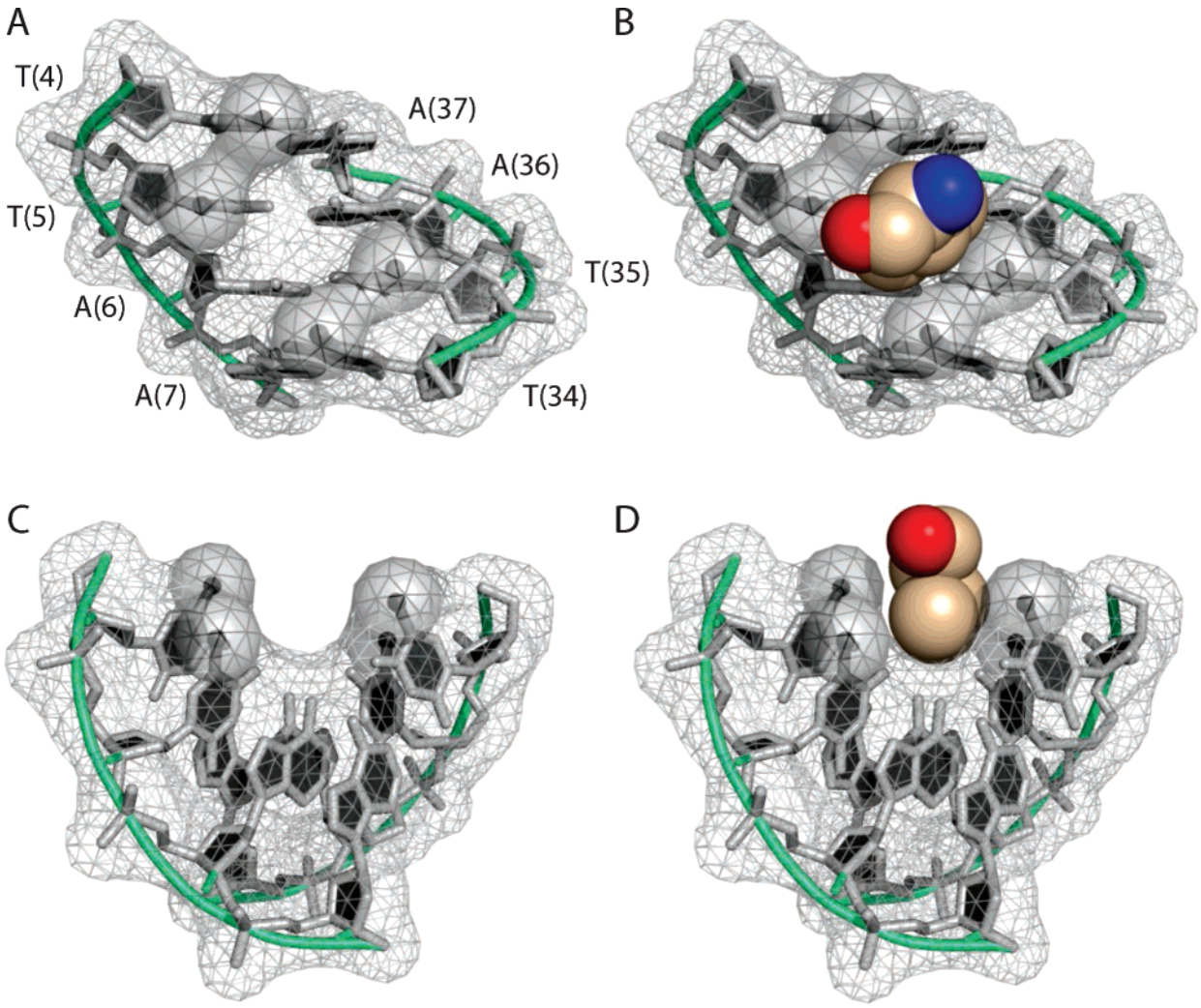

FIGURE 8: Valine cleft. The surface of the DNA atoms composing the R half site valine cleft is represented in the net, and the atoms are in stick representation. A green tube highlights the backbone. The four methyl groups that form the valine cleft are shaded. Val 33 is shown in space filling representation, with C, tan; N, blue; and O, red. (A) DNA only, viewing into the major groove, showing the empty valine cleft. (B) DNA plus Val 33, showing the shape complementarity of Val 33 and valine cleft. (C) Same representation as in panel A except viewed along the major groove. (D) Same representation as in panel B except viewed along the major groove.

within helix 2 are made by Gln21, which contacts the O1P of position 2 [2.8 Å, T(3)] via the backbone nitrogen and the O1P at position 3 via ϵ-NH_2 [2.9 Å, T(4)]. The final side chain to phosphate contact occurs between δ-NH_2 of Asn49 in helix 4 and the O2P at position 7 [2.8 Å, C(33)].

The backbone NH of Ser31 makes a hydrogen bond to the O1P at position 6 [3.0 Å, T(34)], while within helix 4, the backbone NH of Asn46 makes a contact with the O2P at position 7 [2.7 Å, C(33)]. A contact of the backbone NH of Glu44 with the O1P at position 8 [3.0 Å, T(32)] is the only direct protein interaction with DNA^{9T} outside the first seven bases of the operator. This contact, as shown in Figure 7, appears to position the carbonyl of the Glu44 side chain within proximity (<5 Å) of the O2P of thymine at position 8 [T(32)]. A high negative charge density would result from the proximity of phosphates and anionic sidechains. The negative charge may help sequester cations into the minor groove and be a factor in the induction of the B′ state of the noncontacted region.

Protein−Water−DNA. Eighteen water molecules form bridges between the repressor and the DNA^{9T} (Table 2S). Three of these waters link protein side chains to DNA bases, via hydrogen bonds. Ordered water molecules are also associated with the α-helices 1 and 2 of the repressor that do not bridge protein to DNA. The prevalence of these protein-associated ordered solvent molecules decreases toward the center of the complex.

Side Chain−Water−Base. A water molecule (water 24) links the δ-NH_2 of Asn32 and the O4 of thymine at position 3 (Figures 4 and 9A). Both the side chain of Asn32 and the water 24 are defined by a clear electron density and low thermal factors ($B^{W(24)} = 26$ Å^2). The orientation and rotameric state of Asn32 appears to be stabilized by an intramolecular hydrogen bond between the δ-O and the backbone NH. Water 6 mediates the interaction between Nϵ of Gln37 and N7 of adenine at position 5.

A network of interactions between Gln37, Glu42, water molecules, and cytosine at position 7 (Figure 9B) predicts that the base recognition at position 7 is dictated by both residues 37 and 42. The water-mediated contact between Glu42 and N4 cytosine at position 7 is the only observed base contact at this position and may influence the identity of the base pair at that position. We speculate that the replacement of cytosine at position 7 with a purine would likely cause a change in optimal distance and geometry for the water-mediated interaction of Glu42 and position 7,

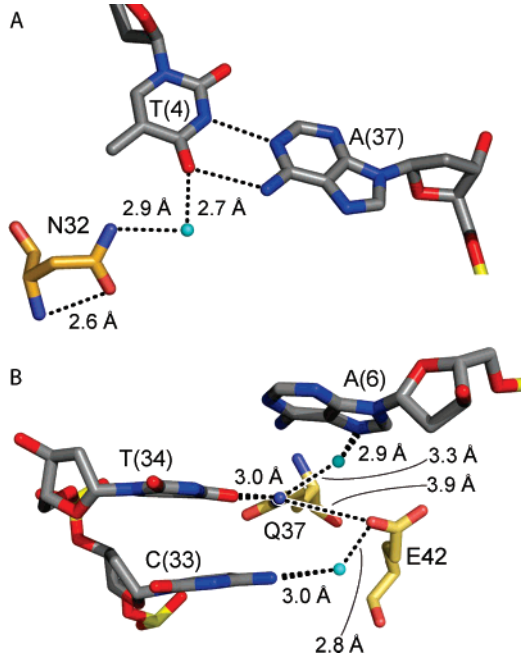

FIGURE 9: Hydrogen bonding interactions that contribute to base recognition. (A) Interaction between Asn32 and O4 of thymine at position 3R is mediated by a water molecule (cyan sphere). (B) Direct hydrogen bond from T(34) to Gln37 is supported by a network of water-mediated interactions from N7 of A(6) and from N(4) of C(33).

resulting in the loss of affinity observed in the solution studies (*31*). However, because the water that mediates the interaction at this position could accept a proton from the N4 cytosine at position 7, as with this structure, or donate a proton to the O4 of thymine, a thymine substitution at this position is tolerated but with a significant decrease in affinity (*31*). The decreased affinity associated with a G-C → A-T substitution at position 7 may arise from a steric clash between the 5-methyl group of a thymine at position 7 and the side chain of Glu37. The steric interference may result in a change in the position of Glu37 and cause a disruption of the hydrogen bond at position 6. Finally, Glu37 and Gln42 may be in close enough proximity (3.9 Å) to interact electrostatically. The removal or change in position of Gln37 would affect the positioning of Glu42 and may result in a less stable water-mediated interaction with cytosine at position 7. Currently, there are no solution studies that have tested the influence of Glu42 on base preference at position 7, and further work will need to be performed to decipher this relationship.

Protein−Protein. All contacts between the two protein monomers are made by helices 4 and 5. Each monomer contains a hydrophobic patch consisting of residues Leu50 and Leu51 of helix 4 and residues Pro61 and Leu65 of helix 5. The hydrophobic patches form the dimer interface. The interface is locked on each side of the hydrophobic patch by salt bridges formed between Lys55 and Asp62 residues located on opposing monomers.

Lattice Contacts. The P22R NTD−DNA^{9T} complex crystallized in the $P4_3$ space group. The DNA fragments are aligned end to end with base pairs of one duplex stacking on the base pairs of the next. The stacking arrangement creates a pseudo-continuous DNA helix. The most disordered region of the DNA occurs at the terminal base pair. The stacking of the DNA strands is facilitated by interactions with a repressor monomer and the operator of a symmetry-related complex.

In addition to the DNA−protein contacts within the complex, there are several lattice contacts, generated by symmetry. The L chain interacts with the major groove. The R chain interacts with the minor groove. Both of these contacts are with the R end of the operator. There are no lattice contacts with the L end of the operator. The asymmetry of these lattice contacts may explain the higher overall thermal factor of the L chain and half site as compared to the R chain of the repressor and R half site.

The L chain of the repressor is positioned over the major groove of a symmetry-related operator, forming a direct contact of ϵ-NH_3 of K18 and O6 of guanine at position 7R (sym1). The R chain is positioned over the minor groove of a symmetry-related operator, allowing a side chain−base contact of ϵ-NH_3 of K18 to N3 of adenine at position 8R (sym2), creating the only break in the primary layer of hydration in the operator. In addition to the side chain−base contacts, several salt bridges link lysines/arginines, and symmetry-related phosphates of the DNA help maintain the crystal lattice.

DISCUSSION

Mechanism of Indirect Readout: Induction of B′-DNA. Here, we define indirect readout as the sensing of a DNA sequence at a distance, in the absence of direct or solvent-mediated protein−base interactions. Haran et al.'s definition of indirect readout as DNA "nearest-neighbor nonadditivity" (*10*) is also a reasonable definition. A universally accepted definition of indirect readout has yet to be established.

The observation of the B′ state in the noncontacted region of DNA^{9T} provides a general predictive model of indirect readout. P22R NTD binds to DNA^{9T} in the major groove, via a helix-turn-helix (HTH) motif. Central noncontacted bases of the operator modulate the affinity of the repressor for the DNA (*23*). The minor groove of this noncontacted region (positions 8R, 9R, 9L, and 8L, sequence 5′-ATAT-3′) is acutely narrowed (Figures 5 and 6). The water molecules within it form a spine of hydration.

DNA is known to be polymorphic, with the ability to adopt a variety of states including A, B, B′ (reviewed by Hud and Plavac (*53*)), and Z-DNA (reviewed by Rich (*54*)). The B′ state of DNA (*55*) differs from the standard B-DNA by (i) a narrow minor groove (*59*, *60*), (ii) a spine of hydration (*61*, *62*) and higher-order hydration within the minor groove (*48*, *63*), (iii) a preference for sequence elements of at least three contiguous ApA or ApT steps (5′ to 3′, TpA steps not favored, G-C base pairs not allowed), (iv) a propensity to bend DNA when appropriately phased (*56*−*58*), (v) monovalent ions within the minor groove (*64*−*68*), (vi) negative propeller twisting (*69*, *70*), (vii) unusual rigidity (*71*, *72*), and (vi) a pre-melting thermal transition that is distinct from duplex melting (*73*−*75*). The B′ helical axis is thought to be linear (*76*, *77*); axial bends arise at the junctions between B′ and flanking B-DNA (*53*). B′ hydration, especially within the minor groove, appears to be cooperative, with many interdependent solvent interactions.

We propose that the transition to the B′ state in the noncontacted region regulates the affinity of P22R for its

binding sites. The stability of the B′ state is modulated by the sequence (ApA, ApT > TpA > CpG, CpC, GpC, GpG) and length. The effect of the sequence of the noncontacted region on the stability of the complex (Figure 1) is consistent with the importance of the B′ state. Sequences that are resistant to the B′ state (*53*) have a low affinity (*23*, *30*). For example, G-C base pairs within the noncontacted region show the greatest destabilizing effect of the P22−operator complex (Figure 1).

The induction of the B′ state appears to be a common feature of DNA−protein complexes. Although not previously described as B′-DNA, we observe narrow minor grooves and elements of spines of hydration in non-A-tract, noncontacted regions of DNA complexes of (i) Narl (1ZG1 (*78*)), (ii) mating-type protein α-1 (1LE8, 1AKH, 1YRN (*79*−*81*)), and (iii) MEF2A (1EGW (*82*)). The induction of the B′ state may allow for optimal contacts between the protein and the operator, favorable hydration, and cation distributions and electrostatics.

Tunnel. In association with P22R NTD, the noncontacted region of the DNA^{9T} forms a tunnel (Figure 7). The solvent appears to be restricted in its translations within the tunnel and could enter and exit only via the openings at the tunnel termini. Our survey of the structural database reveals that minor groove tunnels are a unifying structural theme in complexes between DNA and lambdoid family repressors (*80*), cro dimers, and other HTH proteins (*34*−*36*, *83*−*86*).

Localization of Counterions in the P22 Complex. Biochemical results are consistent with localization of cations within the P22R NTD−DNA^{9T} tunnel and support a role for cations in modulating the DNA−protein stability (*23*, *31*, *32*). The 434 repressor, but not P22R, positions arginine side chains at the lip of the minor groove, directly above the center of the noncontacted region. These cationic side chains contribute to the stability of the 434 repressor−DNA complexes (*5*).

The cationic arginine residues of 434 may be replaced by cationic counterions in the P22R NTD−DNA^{9T} complex. P22R NTD positions two negatively charged glutamate residues (Glu44 and Glu48, Figure 7) in the ceiling of the tunnel, in close proximity to the DNA phosphate groups (4.9 and 7.9 Å, respectively) as well as to each other (4.4 Å). The localization of monovalent cations in the minor groove of the P22R NTD−DNA^{9T} complex is consistent with the induction of the B′ state of DNA^{9T} in the tunnel region. Monovalent cations are known to localize in the minor groove of B′-DNA (*64*−*68*). The tunnel has a high electronegative potential that seems likely to result in cation localization.

Valine Cleft. The structure of the P22R NTD−DNA^{9T} complex suggests that sequence-specificity arises substantially from an interaction of Val33 with a complementary binding cleft on the major groove surface of DNA^{9T} (Figure 8). The cleft is formed by four methyl groups on sequential base pairs (5′-TTAA-3′). This cleft is intrinsic to the DNA sequence and is not induced by P22R NTD binding. The importance of the valine cleft is consistent with the influence of valine 33 on the base pair preferences at positions 3−6 (*30*).

To our knowledge, descriptions of this or related clefts within the major groove are not found in the literature. However, hydrophobic DNA−protein contacts have been noted in structures such as DtxR (*87*), λ cro (*85*) and repressor (*84*), IdeR (*88*, *89*), NarL (*90*), PhoB (*91*), and *E. coli* replication terminator protein (*92*). The importance of hydrophobic interactions in DNA−protein affinity and recognition has been discussed (*87*, *89*, *93*).

TBP versus P22R. A comparison of DNA complexes of TATA box binding protein (*7*, *9*, *94*) and P22R reveals critical differences but also striking similarities. TBP interacts extensively with DNA using nonpolar and hydrophobic interactions and intercalates hydrophobic residues between the base pairs of the DNA. Unlike P22, TBP binds in the minor groove. Unlike P22, TBP induces acute conformational changes in the DNA. However, both P22 and TBP recognize TA base pairs of DNA by shape complementarity. Nonpolar and hydrophobic interactions rather than hydrogen bonds dictate specificity. TBP selects A-T over G-C base pairs via differential deformability and avoids a steric clash with the 2-amino groups of Gs, whereas P22 selects A-T base pairs via interaction with major groove methyl groups via induction of B′-DNA.

Hydrogen Bonding of Gln37 Contributes to the Requirement of an A-T Base Pair at Position 6. A Gln at position 37 is necessary for binding of P22R NTD to the operator (*31*). This requirement is consistent with the interaction between Gln37 and thymine at position 6 (the only direct side chain−base hydrogen bond in the complex, Figures 4 and 9). In a secondary contribution, hydrogen bonding may restrain the side chain of Gln37 in such a way that the aliphatic groups of the side chain contribute hydrophobic interactions with Val33.

Requirement for a G-C Base Pair at Position 7 Is Not Dictated by Simple Directional Interactions. The P22 repressor is intolerant of substitutions at position 7 other than G-C to A-T, which reduces P22R affinity by a factor of 55 (*31*). This G-C requirement at position 7 is highly influenced by Gln37, although no direct contacts link Gln37 with position 7. It appears that subtle solvent organization, shape complementarity, and long-range electrostatic interactions dictate this specificity.

CONCLUSION

We have found that sequence-dependent DNA shape plays a role in direct readout of DNA sequence by the P22 repressor. The binding of P22R NTD to DNA^{9T} appears to be contingent on a complementary cleft for Val33 formed by 5-methyl groups of thymines at positions 3−6.

We have also proposed a model by which a sequence-dependent DNA structure modulates protein−DNA interactions of not only the P22 repressor but many lambdoid family repressors and cro proteins, as well as other HTH family proteins. In our structure of the P22 NTD-9T complex, we find that solvent molecules are organized in a tunnel composed of the dimer interface and the minor groove of the noncontacted region of the operator. We believe that the sequence of bases in this region allows for differences in hydration and affinity for monovalent cations. These differences result in binding affinities that are dependent on the base content in the noncontacted region.

As with most DNA binding proteins, the affinity of the P22 NTD for its operator is not simply accomplished by the recognition of DNA functional groups by protein side chains.

Direct recognition is an important aspect of protein−DNA complex formation. Also contributing are DNA conformation, which is dictated by the sequence, and DNA deformability, which is influenced by sequence and interactions with both protein and solvent. The further categorizing and characterizing of DNA−protein interactions individually as well as how they work in concert with each other is key to understanding and predicting interactions of DNA with proteins and other molecules.

SUPPORTING INFORMATION AVAILABLE

Molecular contacts and helical parameters. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Seeman, N. C., Rosenberg, J. M., and Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins, *Proc. Natl. Acad. Sci. U.S.A. 73*, 804−808.
2. von Hippel, P. H., and Berg, O. G. (1986) On the specificity of DNA−protein interactions, *Proc. Natl. Acad. Sci. U.S.A. 83*, 1608−1612.
3. Zhang, Y., Xi, Z., Hegde, R. S., Shakked, Z., and Crothers, D. M. (2004) Predicting indirect readout effects in protein−DNA interactions, *Proc. Natl. Acad. Sci. U.S.A. 101*, 8337−8341.
4. Anderson, J. E., Ptashne, M., and Harrison, S. C. (1987) Structure of the repressor−operator complex of bacteriophage 434, *Nature (London, U.K.) 326*, 846−852.
5. Koudelka, G. B., Harrison, S. C., and Ptashne, M. (1987) Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and cro, *Nature (London, U.K.) 326*, 886−888.
6. Mauro, S. A., and Koudelka, G. B. (2004) Monovalent cations regulate DNA sequence recognition by 434 repressor, *J. Mol. Biol. 340*, 445−457.
7. Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993) Crystal structure of a yeast TBP/TATA-box complex, *Nature (London, U.K.) 365*, 512−520.
8. DeDecker, B. S., O'Brien, R., Fleming, P. J., Geiger, J. H., Jackson, S. P., and Sigler, P. B. (1996) The crystal structure of a hyperthermophilic archaeal TATA-box binding protein, *J. Mol. Biol. 264*, 1072−1084.
9. Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder, R. G., and Burley, S. K. (1996) Crystal structure of a human TATA box-binding protein/TATA element complex, *Proc. Natl. Acad. Sci. U.S.A. 93*, 4862−4867.
10. Faiger, H., Ivanchenko, M., and Haran, T. E. (2007) Nearest-neighbor non-additivity versus long-range non-additivity in TATA-box structure and its implications for TBP-binding mechanism, *Nucleic Acids Res. 35*, 4409−4419.
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) Crystal structure of trp repressor/operator complex at atomic resolution, *Nature (London, U.K.) 335*, 321−329.
12. Bareket-Samish, A., Cohen, I., and Haran, T. E. (1998) Direct versus indirect readout in the interaction of the trp repressor with non-canonical binding sites, *J. Mol. Biol. 277*, 1071−1080.
13. Somers, W. S., and Phillips, S. E. (1992) Crystal structure of the met repressor−operator complex at 2.8 Å resolution reveals DNA recognition by β-strands, *Nature (London, U.K.) 359*, 387−393.
14. Yang, C. C., and Nash, H. A. (1989) The interaction of *E. coli* IHF protein with its specific binding sites, *Cell 57*, 869−880.
15. Mouw, K. W., and Rice, P. A. (2007) Shaping the *Borrelia burgdorferi* genome: Crystal structure and binding properties of the DNA-bending protein Hbb, *Mol. Microbiol. 63*, 1319−1330.
16. Gabrielsen, O. S., Sentenac, A., and Fromageot, P. (1991) Specific DNA binding by c-Myb: Evidence for a double helix-turn-helix-related motif, *Science (Washington, DC, U.S.) 253*, 1140−1143.
17. Dangi, B., Pelupessey, P., Martin, R. G., Rosner, J. L., Louis, J. M., and Gronenborn, A. M. (2001) Structure and dynamics of MarA−DNA complexes: An NMR investigation, *J. Mol. Biol. 314*, 113−127.
18. Hegde, R. S., Grossman, S. R., Laimins, L. A., and Sigler, P. B. (1992) Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target, *Nature (London, U.K.) 359*, 505−512.
19. Blakaj, D. M., Kattamuri, C., Khrapunov, S., Hegde, R. S., and Brenowitz, M. (2006) Indirect readout of DNA sequence by papillomavirus E2 proteins depends upon net cation uptake, *J. Mol. Biol. 358*, 224−240.
20. Schwabe, J. W., Chapman, L., Finch, J. T., and Rhodes, D. (1993) The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: How receptors discriminate between their response elements, *Cell 75*, 567−578.
21. Chen, S., Vojtechovsky, J., Parkinson, G. N., Ebright, R. H., and Berman, H. M. (2001) Indirect readout of DNA sequence at the primary-kink site in the CAP−DNA complex: DNA binding specificity based on energetics of DNA kinking, *J. Mol. Biol. 314*, 63−74.
22. Joshi, H. K., Etzkorn, C., Chatwell, L., Bitinaite, J., and Horton, N. C. (2006) Alteration of sequence specificity of the type II restriction endonuclease HincII through an indirect readout mechanism, *J. Biol. Chem. 281*, 23852−23869.
23. Wu, L., Vertino, A., and Koudelka, G. B. (1992) Non-contacted bases affect the affinity of synthetic P22 operators for P22 repressor, *J. Biol. Chem. 267*, 9134−9139.
24. Lamoureux, J. S., Maynes, J. T., and Glover, J. N. (2004) Recognition of 5′-YpG-3′ sequences by coupled stacking/hydrogen bonding interactions with amino acid residues, *J. Mol. Biol. 335*, 399−408.
25. Haran, T. E., and Crothers, D. M. (1989) Cooperativity in A-tract structure and bending properties of composite TnAn blocks, *Biochemistry 28*, 2763−2767.
26. Kaiser, A. D. (1957) Mutations in a temperate bacteriophage affecting its ability to lysogenize *Escherichia coli*, *Virology 3*, 42−61.
27. Ptashne, M. (2005) Regulation of transcription: From lambda to eukaryotes, *Trends Biochem. Sci. 30*, 275−279.
28. Maniatis, T., Ptashne, M., Backman, K., Kield, D., Flashman, S., Jeffrey, A., and Maurer, R. (1975) Recognition sequences of repressor and polymerase in the operators of bacteriophage lambda, *Cell 5*, 109−113.
29. Poteete, A. R., Ptashne, M., Ballivet, M., and Eisen, H. (1980) Operator sequences of bacteriophages P22 and 21, *J. Mol. Biol. 137*, 81−91.
30. Hilchey, S. P., and Koudelka, G. B. (1997) DNA-based loss of specificity mutations. Effects of DNA sequence on the contacted and non-contacted base preferences of bacteriophage P22 repressor, *J. Biol. Chem. 272*, 1646−1653.
31. Hilchey, S. P., Wu, L., and Koudelka, G. B. (1997) Recognition of nonconserved bases in the P22 operator by P22 repressor requires specific interactions between repressor and conserved bases, *J. Biol. Chem. 272*, 19898−19905.
32. Wu, L., and Koudelka, G. B. (1993) Sequence-dependent differences in DNA structure influence the affinity of P22 operator for P22 repressor, *J. Biol. Chem. 268*, 18975−18981.
33. Sevilla-Sierra, P., Otting, G., and Wuthrich, K. (1994) Determination of the nuclear magnetic resonance structure of the DNA-binding domain of the P22 c2 repressor (1 to 76) in solution and comparison with the DNA-binding domain of the 434 repressor, *J. Mol. Biol. 235*, 1003−1020.
34. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., and Harrison, S. C. (1988) Recongition of a DNA operator by the repressor of phage 434: A view at high resolution, *Science (Washington, DC, U.S.) 242*, 899−907.
35. Shimon, L. J., and Harrison, S. C. (1993) The phage 434 OR2/R1-69 complex at 2.5 Å resolution, *J. Mol. Biol. 232*, 826−838.
36. Rodgers, D. W., and Harrison, S. C. (1993) The complex between phage 434 repressor DNA-binding domain and operator site OR3: Structural differences between consensus and non-consensus half-sites, *Structure 1*, 227−240.
37. De Anda, J., Poteete, A. R., and Sauer, R. T. (1983) P22 c2 repressor. Domain structure and function, *J. Biol. Chem. 258*, 10536−10542.
38. Amann, E., Brosius, J., and Ptashne, M. (1983) Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in *Escherichia coli*, *Gene 25*, 167−178.
39. Coulondre, C., and Miller, J. H. (1977) Genetic studies of the lac repressor. IV. Mutagenic specificity in the lacI gene of *Escherichia coli*, *J. Mol. Biol. 117*, 577−606.
40. Otwinowski, Z. (1993) *Data Collection and Processing*, pp 56−62, Science and Engineering Research Council, Warrington, U.K.

41. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, *Methods Enzymol. 276*, 307−326.
42. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography and NMR system: A new software suite for macromolecular structure determination, *Acta. Crystallogr., Sect. D: Biol. Crystallogr. 54*, 905−921.
43. Kleywegt, G. J., and Jones, T. A. (1999) Software for handling macromolecular envelopes, *Acta. Crystallogr., Sect. D: Biol. Crystallogr. 55*, 941−944.
44. Ramachandran, G. N., and Sasisekharan, V. (1968) Conformation of polypeptides and proteins, *Adv. Protein Chem. 23*, 283−438.
45. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) Procheck: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
46. Wharton, R. P., and Ptashne, M. (1985) Changing the binding specificity of a repressor by redesigning an α-helix, *Nature (London, U.K.) 316*, 601−605.
47. Mack, D. R., Chiu, T. K., and Dickerson, R. E. (2001) Intrinsic bending and deformability at the T-A step of CCTTTAAAGG: A comparative analysis of T-A and A-T steps within A-tracts, *J. Mol. Biol. 312*, 1037−1049.
48. Shui, X., Sines, C., McFail-Isom, L., VanDerveer, D., and Williams, L. D. (1998) Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations, *Biochemistry 37*, 16877−16887.
49. Williams, L. D. (2005) Between objectivity and whim: Nucleic acid structural biology, in *Topics in Current Chemistry* (Chaires, J. B., and Waring, M., Eds.), pp 77−88, Springer, Heidelberg.
50. Howerton, S. B., Sines, C. C., VanDerveer, D., and Williams, L. D. (2001) Locating monovalent cations in the grooves of B-DNA, *Biochemistry 40*, 10023−10031.
51. McFail-Isom, L., Sines, C., and Williams, L. D. (1999) DNA structure: Cations in charge? *Curr. Opin. Struct. Biol. 9*, 298−304.
52. Poteete, A. R., and Ptashne, M. (1982) Control of transcription by the bacteriophage P22 repressor, *J. Mol. Biol. 157*, 21−48.
53. Hud, N. V., and Plavec, J. (2003) A unified model for the origin of DNA sequence-directed curvature, *Biopolymers 69*, 144−158.
54. Rich, A. (1993) DNA comes in many forms, *Gene 135*, 99−109.
55. Arnott, S., and Hukins, D. W. L. (1972) Optimized parameters for A-DNA and B-DNA, *Biochem. Biophys. Res. Commun. 47*, 1504−1509.
56. Marini, J. C., Levene, S. D., Crothers, D. M., and Englund, P. T. (1982) Bent helical structure in kinetoplast DNA, *Proc. Natl. Acad. Sci. U.S.A. 79*, 7664−7668.
57. Wu, H.-M., and Crothers, D. M. (1984) The locus of sequence-directed and protein-induced DNA bending, *Nature (London, U.K.) 308*, 509−513.
58. Zinkel, S. S., and Crothers, D. M. (1987) DNA bend direction by phase sensitive detection, *Nature (London, U.K.) 328*, 178−181.
59. Alexeev, D. G., Lipanov, A. A., and Skuratovskii, I. Y. (1987) Poly(dA)-poly(dT) is a B-type double helix with a distinctively narrow minor groove, *Nature (London, U.K.) 325*, 821−823.
60. Burkhoff, A. M., and Tullius, T. D. (1987) The unusual conformation adopted by the adenine tracts in kinetoplast DNA, *Cell 48*, 935−943.
61. Kopka, M. L., Fratini, A. V., Drew, H. R., and Dickerson, R. E. (1983) Ordered water structure around a B-DNA dodecamer. A quantitative study, *J. Mol. Biol. 163*, 129−146.
62. Drew, H. R., and Dickerson, R. E. (1981) Structure of a B-DNA dodecamer. III. Geometry of hydration, *J. Mol. Biol. 151*, 535−556.
63. Woods, K. K., Maehigashi, T., Howerton, S. B., Sines, C. C., Tannenbaum, S., and Williams, L. D. (2004) High-resolution structure of an extended A-tract: $[d(CGCAAATTTGCG)]_2$, *J. Am. Chem. Soc. 126*, 15330−15331.
64. Woods, K., McFail-Isom, L., Sines, C. C., Howerton, S. B., Stephens, R. K., and Williams, L. D. (2000) Monovalent cations sequester within the A-tract minor groove of $[d(CGCGAATTCGCG)]_2$, *J. Am. Chem. Soc. 122*, 1546−1547.
65. Shui, X., McFail-Isom, L., Hu, G. G., and Williams, L. D. (1998) The B-DNA dodecamer at high resolution reveals a spine of water on sodium, *Biochemistry 37*, 8341−8355.
66. Tereshko, V., Minasov, G., and Egli, M. (1999) A "hydrat-ion" spine in a B-DNA minor groove, *J. Am. Chem. Soc. 121*, 3590−3595.
67. Hud, N. V., Sklenar, V., and Feigon, J. (1999) Localization of ammonium ions in the minor groove of DNA duplexes in solution and the origin of DNA A-tract bending, *J. Mol. Biol. 286*, 651−660.
68. Cesare Marincola, F., Denisov, V. P., and Halle, B. (2004) Competitive Na(+) and Rb(+) binding in the minor groove of DNA, *J. Am. Chem. Soc. 126*, 6739−6750.
69. Wing, R., Drew, H., Takano, T., Broka, C., Takana, S., Itakura, K., and Dickerson, R. E. (1980) Crystal structure analysis of a complete turn of B-DNA, *Nature (London, U.K.) 287*, 755−758.
70. Dickerson, R. E., and Drew, H. R. (1981) Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure, *J. Mol. Biol. 149*, 761−786.
71. Rhodes, D. (1979) Nucleosome cores reconstituted from poly (dA-dT) and the octamer of histones, *Nucleic Acids Res. 6*, 1805−1816.
72. Simpson, R. T., and Kunzler, P. (1979) Chromatin and core particles formed from the inner histones and synthetic polydeoxyribonucleotides of defined sequence, *Nucleic Acids Res. 6*, 1387−1415.
73. Herrera, J. E., and Chaires, J. B. (1989) A premelting conformational transition in poly(dA)-poly(dT) coupled to daunomycin binding, *Biochemistry 28*, 1993−2000.
74. Chan, S. S., Breslauer, K. J., Austin, R. H., and Hogan, M. E. (1993) Thermodynamics and premelting conformational changes of phased $(dA)_5$ tracts, *Biochemistry 32*, 11776−11784.
75. Augustyn, K. E., Wojtuszewski, K., Hawkins, M. E., Knutson, J. R., and Mukerji, I. (2006) Examination of the premelting transition of DNA A-tracts using a fluorescent adenosine analogue, *Biochemistry 45*, 5039−5047.
76. Nelson, H. C. M., Finch, J. T., Luisi, B. F., and Klug, A. (1987) The structure of an oligo(dA)·oligo(dT) tract and its biological implications, *Nature (London, U.K.) 330*, 221−226.
77. Coll, M., Frederick, C. A., Wang, A. H.-J., and Rich, A. (1987) A bifurcated hydrogen-bonded conformation in the d(A-T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin, *Proc. Natl. Acad. Sci. U.S.A. 84*, 8385−8389.
78. Baikalov, I., Schroder, I., Kaczor-Grzeskowiak, M., Cascio, D., Gunsalus, R. P., and Dickerson, R. E. (1998) NarL dimerization? Suggestive evidence from a new crystal form, *Biochemistry 37*, 3665−3676.
79. Ke, A., Mathias, J. R., Vershon, A. K., and Wolberger, C. (2002) Structural and thermodynamic characterization of the DNA binding properties of a triple alanine mutant of MATalpha2, *Structure 10*, 961−971.
80. Li, T., Jin, Y., Vershon, A. K., and Wolberger, C. (1998) Crystal structure of the MATa1/MATalpha2 homeodomain heterodimer in complex with DNA containing an A-tract, *Nucleic Acids Res. 26*, 5707−5718.
81. Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995) Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA, *Science (Washington, DC, U.S.) 270*, 262−269.
82. Santelli, E., and Richmond, T. J. (2000) Crystal structure of MEF2A core bound to DNA at 1.5 Å resolution, *J. Mol. Biol. 297*, 437−449.
83. Mondragon, A., and Harrison, S. C. (1991) The phage 434 Cro/OR1 complex at 2.5 Å resolution, *J. Mol. Biol. 219*, 321−334.
84. Beamer, L. J., and Pabo, C. O. (1992) Refined 1.8 Å crystal structure of the lambda repressor−operator complex, *J. Mol. Biol. 227*, 177−196.
85. Albright, R. A., and Matthews, B. W. (1998) Crystal structure of lambda-Cro bound to a consensus operator at 3.0 Å resolution, *J. Mol. Biol. 280*, 137−151.
86. Jain, D., Kim, Y., Maxwell, K. L., Beasley, S., Zhang, R., Gussin, G. N., Edwards, A. M., and Darst, S. A. (2005) Crystal structure of bacteriophage lambda cII and its DNA complex, *Mol. Cell 19*, 259−269.
87. Chen, C. S., White, A., Love, J., Murphy, J. R., and Ringe, D. (2000) Methyl groups of thymine bases are important for nucleic acid recognition by DtxR, *Biochemistry 39*, 10397−10407.
88. Wisedchaisri, G., Chou, C. J., Wu, M., Roach, C., Rice, A. E., Holmes, R. K., Beeson, C., and Hol, W. G. (2007) Crystal structures, metal activation, and DNA-binding properties of two-domain IdeR from *Mycobacterium tuberculosis*, *Biochemistry. 46*, 436−447.

89. Wisedchaisri, G., Holmes, R. K., and Hol, W. G. (2004) Crystal structure of an IdeR−DNA complex reveals a conformational change in activated IdeR for base-specific interactions, *J. Mol. Biol. 342*, 1155−1169.
90. Maris, A. E., Kaczor-Grzeskowiak, M., Ma, Z., Kopka, M. L., Gunsalus, R. P., and Dickerson, R. E. (2005) Primary and secondary modes of DNA recognition by the NarL two-component response regulator, *Biochemistry 44*, 14538−14552.
91. Blanco, A. G., Sola, M., Gomis-Ruth, F. X., and Coll, M. (2002) Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator, *Structure 10*, 701−713.
92. Kamada, K., Horiuchi, T., Ohsumi, K., Shimamoto, N., and Morikawa, K. (1996) Structure of a replication−terminator protein complexed with DNA, *Nature (London, U.K.) 383*, 598−603.
93. Oobatake, M., Kono, H., Wang, Y., and Sarai, A. (2003) Anatomy of specific interactions between lambda repressor and operator DNA, *Proteins 53*, 33−43.
94. Burley, S. K. (1996) The TATA box binding protein, *Curr. Opin. Struct. Biol. 6*, 69−75.

BI701826F