

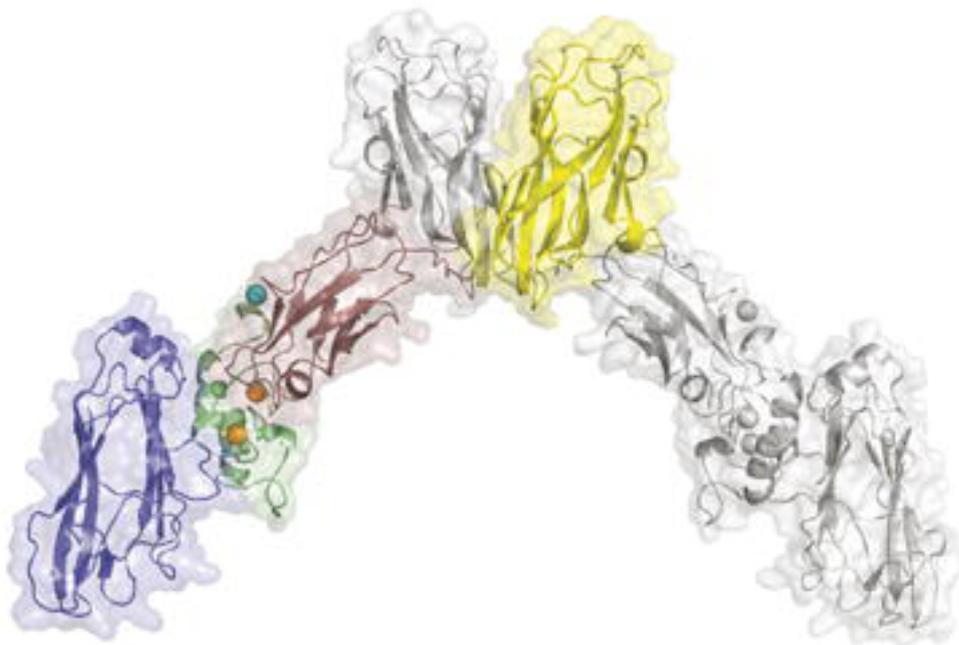


Volume 396, Issue 4, 5 March 2010

ISSN 0022-2836

# JMIB

JOURNAL OF MOLECULAR BIOLOGY



0022-2836(20100305)396:4;1-E

**JMB**Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

**ScienceDirect**


# Sequence Recognition of DNA by Protein-Induced Conformational Transitions

Derrick Watkins<sup>1</sup>, Srividya Mohan<sup>1</sup>, Gerald B. Koudelka<sup>2\*</sup>  
and Loren Dean Williams<sup>1\*</sup>

<sup>1</sup>*School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA*

<sup>2</sup>*Department of Biological Sciences, 607 Cooke Hall, State University of New York at Buffalo, Buffalo, NY 14260, USA*

Received 14 October 2009;  
received in revised form  
23 December 2009;  
accepted 25 December 2009  
Available online  
4 January 2010

The binding of proteins to specific sequences of DNA is an important feature of virtually all DNA transactions. Proteins recognize specific DNA sequences using both direct readout (sensing types and positions of DNA functional groups) and indirect readout (sensing DNA conformation and deformability). Previously we showed that the P22 c2 repressor N-terminal domain (P22R NTD) forces the central non-contacted 5'-ATAT-3' sequence of the DNA operator into the B' state, a state known to affect DNA hydration, rigidity and bending. Usually the B' state, with a narrow minor groove and a spine of hydration, is reserved for A-tract DNA (TpA steps disrupt A-tracts). Here, we have co-crystallized P22R NTD with an operator containing a central 5'-ACGT-3' sequence in the non-contacted region. C·G base pairs have not previously been observed in the B' state and are thought to prevent it. However, P22R NTD induces a narrow minor groove and a spine of hydration to 5'-ACGT-3'. We observe that C·G base pairs have distinctive destabilizing and disordering effects on the spine of hydration. It appears that the reduced stability of the spine results in a higher energy cost for the B to B' transition. The differential effect of DNA sequence on the barrier to this transition allows the protein to sense the non-contacted DNA sequence.

© 2009 Published by Elsevier Ltd.

Edited by K. Morikawa

**Keywords:** indirect readout; minor groove; P22 repressor; spine of hydration; A-tract DNA

## Introduction

The binding of proteins to specific sequences of DNA is an important feature of virtually all DNA transactions. Here we demonstrate the mechanism by which the affinity of a protein for DNA is attenuated by sequence-dependent DNA conformation and deformability.

## Protein recognition of DNA

When recognizing DNAs of specific sequence, proteins use both direct and indirect readout. In direct readout, DNA bases make van der Waals contacts with protein side chains and backbone. Direct readout can employ hydrogen bonding in

analogy with the proposal by Seeman *et al.*<sup>1</sup> Direct readout can also employ shape complementarity in the absence of hydrogen bonding.<sup>2</sup>

In indirect readout, DNA sequence is detected by its propensity to assume noncanonical conformations. DNA can adopt a variety of states ranging from A to B and B' (reviewed by Hud and Plavec<sup>3</sup>), in addition to Z (reviewed by Rich<sup>4</sup>). The propensity for DNA to adopt various conformational states is modulated by sequence. Proteins use DNA conformational polymorphism and deformability as recognition signals. The roles of DNA conformation and deformability in protein recognition was put in a quantitative framework by Olson *et al.*<sup>5</sup> McFail-Isom *et al.* have described how DNA electrostatic potential and charged species can alter DNA conformation, especially minor groove width.<sup>6</sup> Koudelka and Carlson<sup>7</sup> and Zhang *et al.*<sup>8</sup> experimentally determined relationships between DNA flexibility and protein affinity. More recently, Rohs *et al.* have proposed that electrostatic potential and minor groove width are recognized by proteins.<sup>9</sup>

\*Corresponding authors. E-mail addresses:  
[koudelka@buffalo.edu](mailto:koudelka@buffalo.edu); [ldw@gatech.edu](mailto:ldw@gatech.edu).

Abbreviations used: P22R NTD, P22 c2 repressor N-terminal domain; ESP, electrostatic surface potential; PEG, polyethylene glycol; MPD, methyl-2,4-pentanediol.

The previous X-ray structure of the P22 c2 repressor N-terminal domain (P22R NTD) in a complex with a DNA operator<sup>2</sup> confirms that direct readout of helix–turn–helix proteins involves interactions of recognition helices with the floor of the major groove (Fig. 1). The observed shape complementarity between the DNA and the P22R recognition helix (Fig. 2) is consistent with effects of mutations on P22R–DNA complex stability.<sup>11,12</sup> In this case, recognition is achieved primarily by shape complementarity rather than hydrogen-bonding interactions.

### A molecular-level model of indirect readout

Solution experiments demonstrated indirect readout of the non-contacted bases at the center of the P22R DNA operator.<sup>10,12,13</sup> The sequence of the central non-contacted bases modulates the affinity of P22R. The P22R NTD–DNA structure<sup>2</sup> suggested a molecular-level mechanistic model for this indirect readout. The three-dimensional structure indicates that P22R induces a transition of the non-contacted region from B-DNA to B'-DNA.<sup>2</sup> In the P22R NTD–DNA complex, the minor groove is narrow and is hydrated by an "A-tract-like" zigzag spine, despite the absence of an A-tract DNA sequence. The structure led us to propose a model of indirect readout in which P22R induces the B' transition of the non-contacted region of the DNA. The differential effect of DNA sequence on the barrier to this transition allows the protein to sense the non-contacted DNA sequence.

### The B' state

The B' state of DNA<sup>14</sup> (Figs. 3 and 4) was first described by Arnott and Hukins<sup>15</sup> and has been shown by Tullius and coworkers<sup>16–18</sup> to form in A-tracts, which are regions of contiguous 5'-ApA-3' and/or ApT steps. TpA steps provide a barrier to formation of the B' state. That barrier can be overcome: the B' state can be induced by proteins<sup>2</sup> or crystal lattice effects.<sup>19</sup> Tevis *et al.* have data suggesting that certain minor groove binders can induce the B' state in TpA steps.<sup>20</sup> However, C·G base pairs appear to provide the largest barrier to the transition and have not been observed previously in the B' state.

The B' state differs from the standard B-DNA by (i) a narrow minor groove (Fig. 3a),<sup>16,21</sup> (ii) a spine of hydration<sup>22,23</sup> and higher-order hydration within the minor groove<sup>24–26</sup> (Fig. 3b), (iii) a propensity to bend DNA when appropriately phased,<sup>27–29</sup> (iv) monovalent cations ions within the minor groove,<sup>25,30–33</sup> (v) negative propeller twisting of base pairs,<sup>34,35</sup> (vi) unusual rigidity,<sup>36,37</sup> and (vii) a "premelting" thermal transition that is distinct from duplex melting.<sup>38–40</sup> The B' helical axis is thought to be linear.<sup>41,42</sup> Axial bends arise at the junctions between B' and flanking B-DNA.<sup>3</sup> The hydration of B'-DNA within the minor groove appears to be monolithic

and cooperative, resulting in an interdependent hydration assembly (Fig. 3).

### Testing the model

Our model of indirect readout predicts that within the central region of the operator, binding by P22R would force any sequence, even sequences containing C·G base pairs, into the B' state. The non-contacted sequence is "read" by the ease with which it adopts the B' state. To test the model, we determined the X-ray structure of the P22R NTD in complex with the synthetic operator DNA<sup>9C</sup>. This complex contains a CpG step at the center of the non-contacted region of the DNA operator. The model predicts that this complex should reveal, for the first time, C·G base pairs in the B' state. Indeed, we observe that the non-contacted region of the complex has been forced into the B' state. The structure helps explain the origin of the sequence dependence of B'-DNA stability. The hydration of a CpG step, even when forced into the B' state, differs from that of ApA, ApT or TpA steps, and appears to destabilize the B' state. Finally, we crystallized the DNA<sup>9C</sup> and DNA<sup>9T</sup> complexes with P22R NTD in the presence of Tl<sup>+</sup> or Rb<sup>+</sup>. Both Tl<sup>+</sup> and Rb<sup>+</sup> have been shown to mimic K<sup>+</sup> in interactions with DNA, RNA and proteins.<sup>43–46</sup> Tl<sup>+</sup> and Rb<sup>+</sup> are directly observable in crystallographic studies. Thus, Tl<sup>+</sup> and Rb<sup>+</sup> can reveal the positions at which cations bind in the P22R–DNA complexes without resorting to analysis of subtle differences in coordination chemistry.

## Results

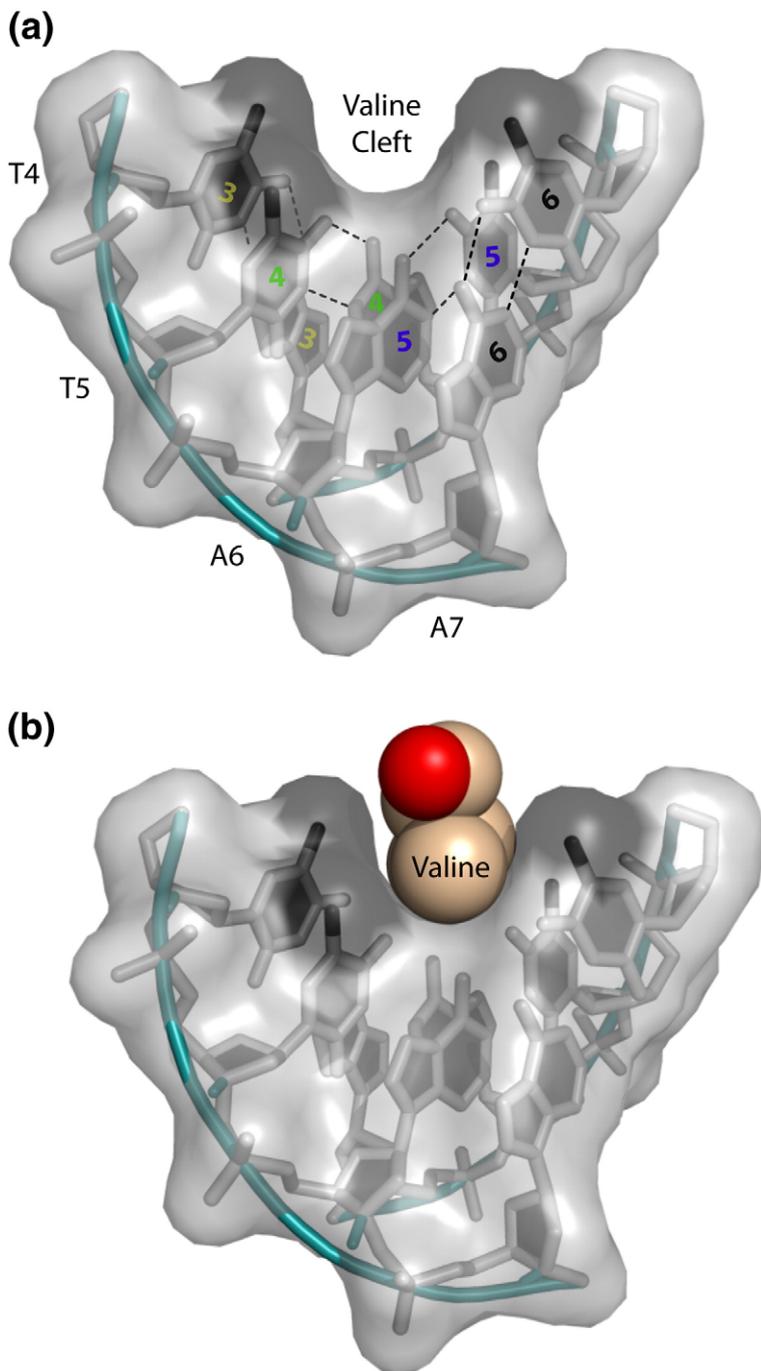
### The global complex

The maps surrounding the DNA<sup>9C</sup> and P22R NTD are clean and continuous (Fig. 5). The global architecture of the P22R NTD–DNA<sup>9C</sup> complex is unaffected by a change from TpA to CpG in the non-contacted region. The helical parameters confirm that the overall conformations of DNA<sup>9C</sup> and DNA<sup>9T</sup> are conserved. The DNA–protein contacts in the two complexes are virtually identical.

### Direct readout: the valine cleft

The Val33 interaction with the valine cleft on the DNA is a feature of P22R NTD complexes and has not been observed in other DNA–protein complexes. Val33 is inserted into a prominent cleft in the major groove of both DNA<sup>9C</sup> and DNA<sup>9T</sup> (Fig. 2). The cleft is formed by the methyl groups of four thymidines. The shape complementarity of the DNA surface with that of the protein locks the position of the DNA relative to the protein. The interactions of Val33 within the pocket make a primary contribution to DNA sequence specificity. The position Val33 is nearly identical in the DNA<sup>9C</sup> and DNA<sup>9T</sup> complexes, with the atoms





**Fig. 2.** The valine cleft. (a) The 5-TTAA-3' sequence from the contacted region of the DNA (positions 3R–6R or 3L–6L) viewing along the major groove, showing the “empty” valine cleft. The four methyl groups that form the walls of the valine cleft are darkened. Operator positions are indicated on the bases. Watson–Crick hydrogen bonds are shown by dashed lines. (b) The DNA plus valine 33, showing the shape complementarity of the valine and the valine cleft.

involved in the cleft superimposing with an RMSD of 0.086 Å.

#### Indirect readout: the noncontacted region

P22R NTD does not contact the DNA bases in the central region of the operator. The central non-contacted bases influence the affinity of the protein for the DNA.

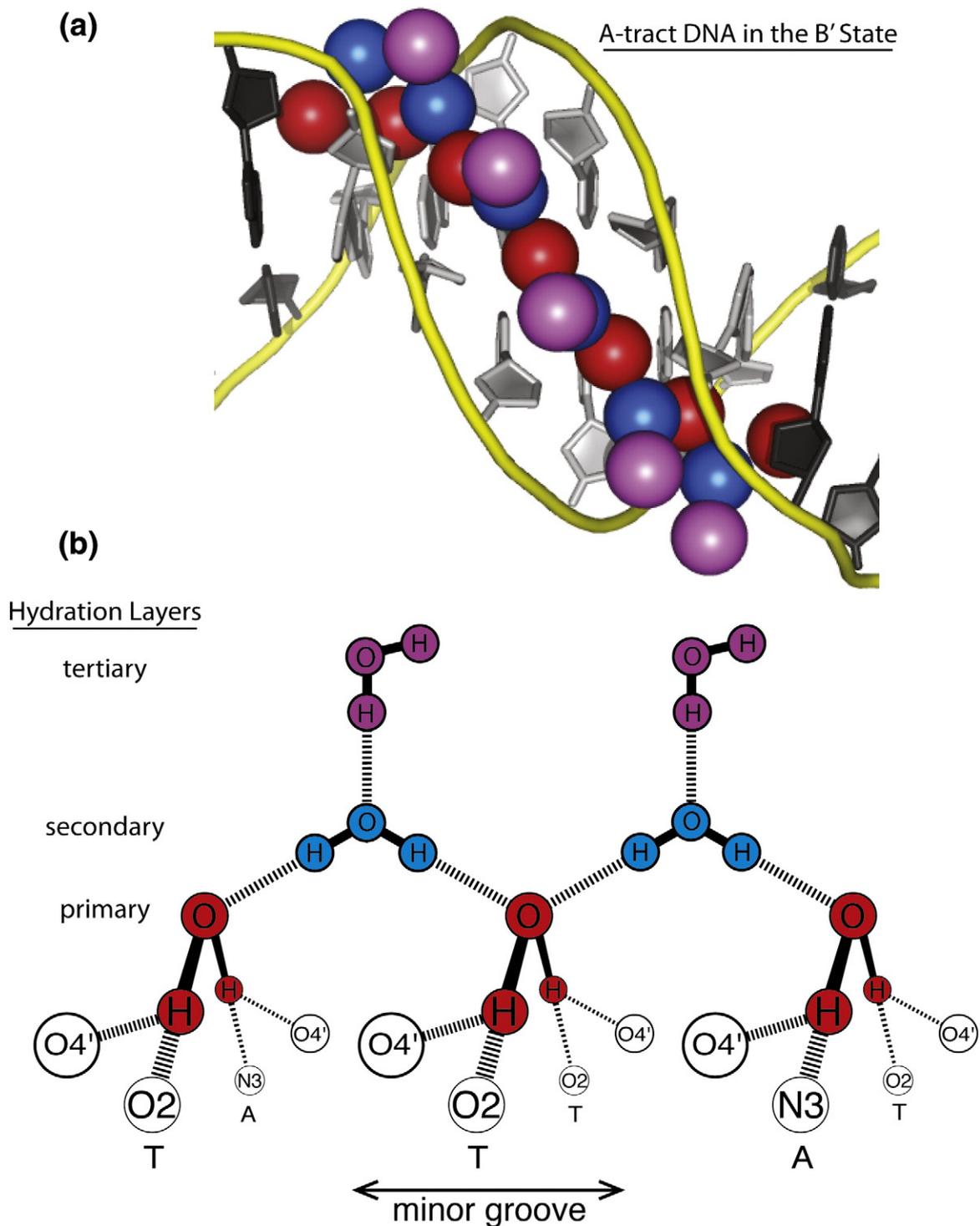
#### Minor groove width

The minor groove at the center of the P22R NTD–DNA<sup>9C</sup> complex is narrow (Fig. 4). The

groove width reaches a minimum at the center of the non-contacted region, at positions 9L and 9R. The groove begins to widen going away from the center (8R through 5R and 8L through 5L). The minor groove then narrows again in the A-tracts at each end of the operator (4R through 1R and 4L through 1L). The minor groove width profiles are the same in P22R NTD–DNA<sup>9C</sup> and P22R NTD–DNA<sup>9T</sup>.

#### The induced B' state

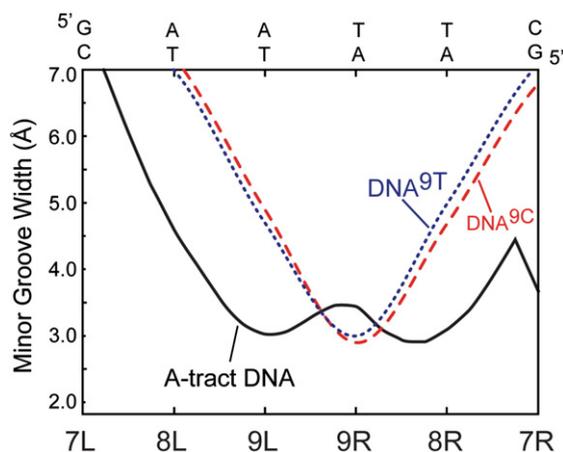
In spite of differences in sequence, the conformations of the DNA in the non-contacted regions of



**Fig. 3.** A-tract DNA in the B' state. (a) The minor groove of B' DNA is narrow and is occupied by a zigzag spine, with additional ordered layers of hydration. The water molecules of the primary layer of the spine are red. The secondary layer is blue. The tertiary layer is purple. The quaternary layer is not shown. This figure is made from PDB entry 1S2R. (b) In A-tract DNA, hydrogen-bond acceptors localize water molecules, restraining both translations and rotations. In the absence of C-G base pairs, the minor groove lacks hydrogen-bond donors. Water molecules in the primary layer form bifurcated hydrogen bonds with the DNA, in which a given proton is shared by two acceptors. Each primary water molecule donates each proton to two hydrogen-bond acceptors.

DNA<sup>9C</sup> and DNA<sup>9T</sup> complexes are similar to each other and to that of A-tract DNA. The RMSD of backbone atoms between the non-contacted region

of the DNA<sup>9C</sup> and DNA<sup>9T</sup> is only 0.29 Å. As indicated by the narrow minor groove (Fig. 4) and a spine-like pattern of hydration (below), the non-



**Fig. 4.** Minor groove profiles. The minor grooves of both DNA<sup>9T</sup> (blue) and DNA<sup>9C</sup> (red) are narrow, as is the groove of A-tract DNA alone (black; PDB entry 355D). The sequence along the bottom axis indicates the positions of DNA<sup>9T</sup> and DNA<sup>9C</sup>, as illustrated in Fig. 1. The sequence along the top axis is that of the A-tract.

contacted region of the DNA<sup>9C</sup>, like that of the non-contacted region in the DNA<sup>9T</sup> complex, is in the B' state.

The B' state is induced by P22R NTD. DNA<sup>9C</sup> in the complex with P22R NTD represents, to our knowledge, the first observation of a C:G base pair in the B' state. For example, several fragments of unbound DNA containing the sequence element ACGT are found in the Protein Data Bank (PDB) [entries 424D (B-form), 1HQ7 (B-form), 117D (A-form), 116D (A-form)]. None of these adopt the B' state. A large and consistent body of literature on the sequence dependence of DNA conformation in solution indicates that unbound DNA containing C:G base pairs does not adopt the B' state.

#### Protein–phosphate contacts and electrostatic potentials

The interactions of the P22R NTD with DNA<sup>9C</sup> and DNA<sup>9T</sup> in the non-contacted region are restricted to contacts with phosphate backbone. The  $\delta$ -NH<sub>2</sub> of Asn49 of the L monomer in the DNA<sup>9C</sup> complex contacts the O2P at position 7 [2.8 Å, C(33)]. The backbone NH of Asn46 contacts the O2P at position 7 [2.7 Å, C(33)]. The N<sup>ε</sup> of Trp38 contacts the O1P at position 7 [2.8 Å, C(33)]. The backbone NH of Glu44 contacts the O1P at position 8 [3.0 Å, T(32)]. The position of the carbonyl of the Glu44 side chains remains consistent with those of the DNA<sup>9T</sup> complex and are within proximity (<5 Å) of the O2P of thymine at position 8 [T(32)]. These phosphate contacts are mirrored by contacts of the R monomer of the repressor dimer in both location and contact distances. These contacts between the repressor and the phosphate backbone are dependent on the DNA having a narrow minor groove in the non-contacted region.

Consistent with the idea that protein binding forces the central minor groove to narrow and induces the B' state, the results of electrostatic surface potential (ESP) calculations demonstrate that the electrostatic surface of the protein is complementary to that of the B' state. Specifically, the calculations indicate that the repressor contains a strip of positive ESP running approximately diagonally across the DNA binding surface of the repressor. This strip of positive potential contains four discrete regions of high positive ESP (Fig. 6). These four regions create “phosphate pockets.” The two most central pockets form a complementary electrostatic surface that can bind the phosphates distributed across the two DNA strands of the central non-contacted region only when that region is in the B' state.

ESP calculations also reveal that the most electro-positive ( $\sim 8$  kT/e<sup>-</sup>) region on the surface of the P22R NTD occurs at the outer edges of the protein. These regions contact the phosphates and define the ends of the DNA binding site.

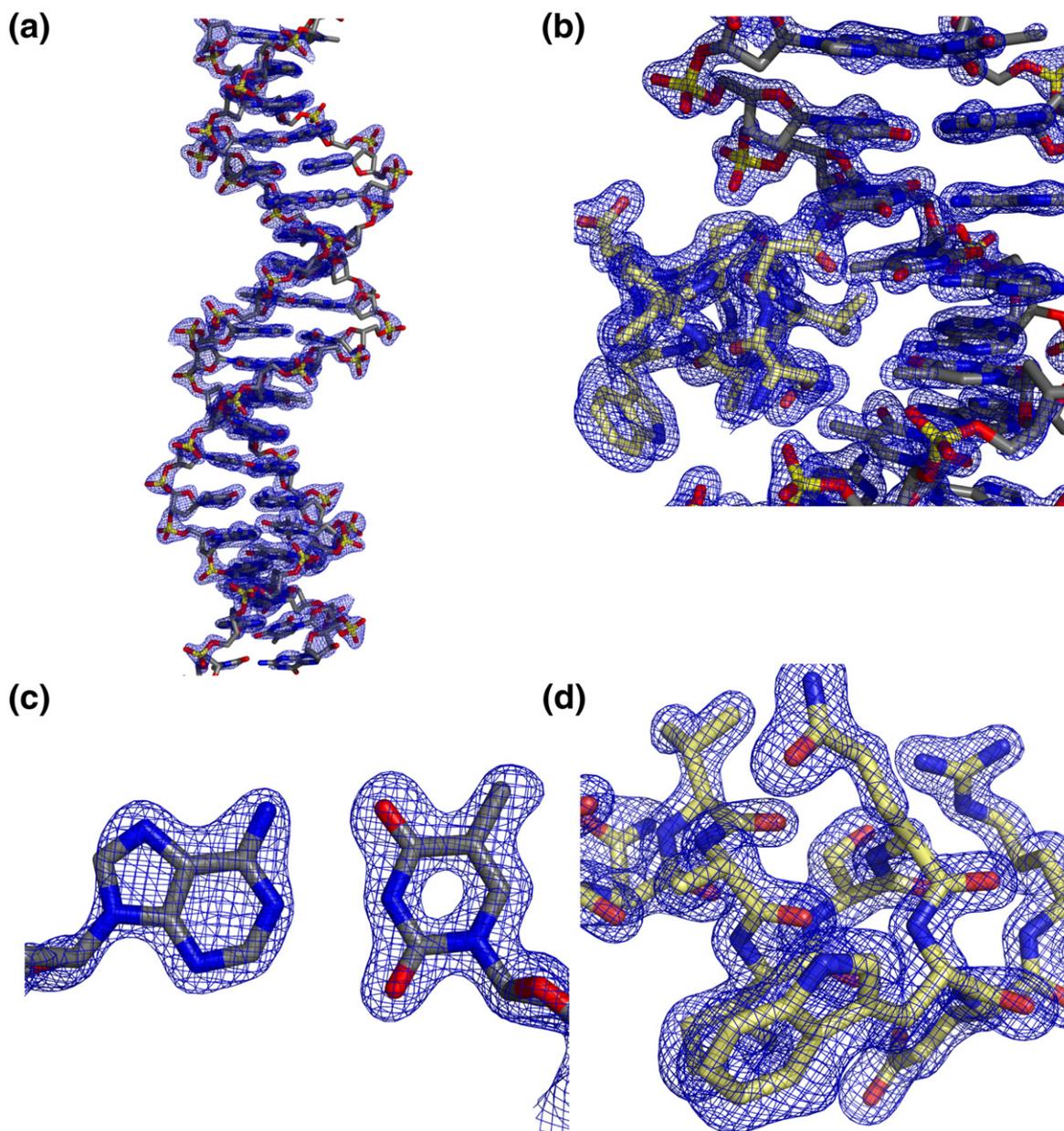
#### The tunnel

At the center of the binding site, in the non-contacted region, the DNA minor groove and the repressor combine to form a tunnel (Fig. 1). Access to the central minor groove requires entry through the ends of the tunnel. The structure of the tunnel is nearly identical between the complexes of P22R NTD with DNA<sup>9C</sup> and DNA<sup>9T</sup>. For example, the conformations of the walls of the tunnel in the DNA<sup>9C</sup> structure, made up of the sugar phosphate backbone, are unperturbed relative to their position in the DNA<sup>9T</sup> structure. The ceiling of the tunnel in both complexes is formed by the dimer interface of the repressor, which is made up from the loop region connecting helices 3 and 4 and the N-terminal region of helix 4.

#### The floor of the tunnel

The positions of hydrogen-bond acceptors on the floor and walls of the minor groove are conserved in the DNA<sup>9T</sup> (Fig. 7) and DNA<sup>9C</sup> (Fig. 8) complexes. The N3 atoms of the purines and the O2 atoms of the pyrimidines at positions 9R and 9L are invariant. However, the DNA<sup>9C</sup> complex, but not the DNA<sup>9T</sup> complex, contains hydrogen-bond donors. These 2-amino groups (N2 atoms) of the guanine residues of the non-contacted region of DNA<sup>9C</sup> impact the electrostatics, hydrogen bonding, and shape characteristics of the tunnel floor. The 2-amino group of guanine is relatively electropositive, is a hydrogen-bond donor, and protrudes from the surface out into the minor groove.

In the non-contacted tunnel region of the DNA<sup>9T</sup> operator, the floor of the minor groove is at a greater negative potential than in the A-tracts of the contacted region (positions 1R–4R and 1L–4L). Thus, the induced B' state has a lower potential ( $-20$  kT/e<sup>-</sup>) than the intrinsic B' state ( $-14$  kT/e<sup>-</sup>).



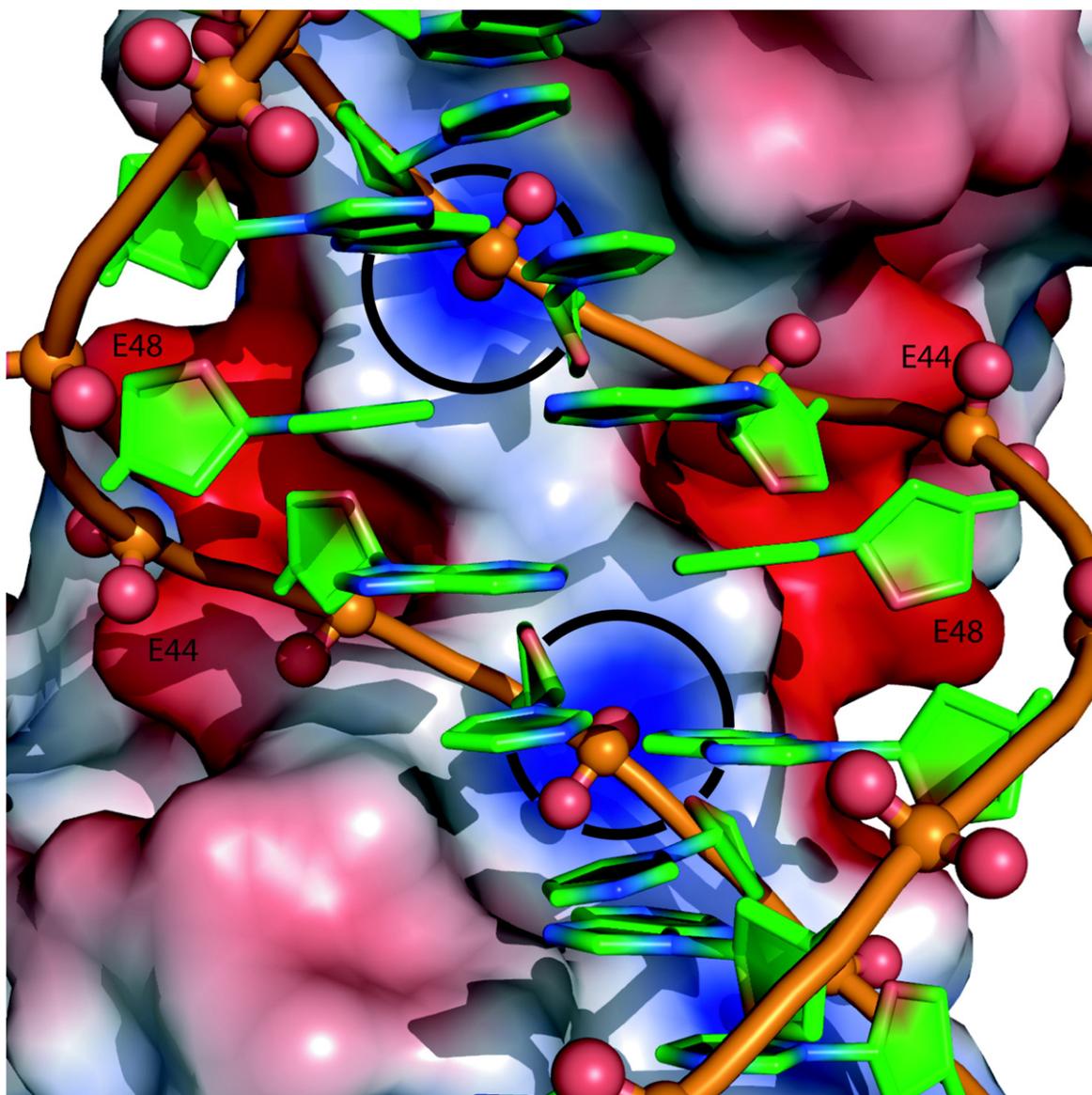
**Fig. 5.** Electron density maps of protein and DNA. The electron density defining the DNA, protein and solvent is of high quality. Shown here is the final sum ( $12F_o - F_c$ ) electron density contoured at  $1.5\sigma$ . (a) Electron density surrounding the DNA duplex. (b) Electron density surrounding the DNA and the R-subunit recognition helix. (c) Electron density surrounding base pair T(9)–A(32) (position 8R). (d) Electron density surrounding the recognition helix (residues R31–R40).

The ESP calculated for the complex indicates that the protein contributes to the negative potential of the tunnel both by forcing the groove to narrow and from the proximity of negatively charged side chains.

**Minor groove hydration.** The introduction of the 2-amino groups of guanine into the central minor groove of the P22R–DNA complex (i.e., the switch from DNA<sup>9T</sup> to DNA<sup>9C</sup>) has profound effects on positions, stability and mobility of the solvent within the groove. In DNA<sup>9T</sup>, each water molecule near the floor of the tunnel interacts with a well-defined set of hydrogen-bond acceptors, occupying

a single highly localized binding site (Fig. 7). The position of each relevant water molecule of DNA<sup>9T</sup> is defined by a spherical peak of electron density (Fig. 7b). The distances between the water molecules of DNA<sup>9T</sup> fall within a narrow range ( $3.05 \pm 0.37$  Å). Their thermal factors are on the same order as those of the bases.

In contrast, the corresponding water molecules of DNA<sup>9C</sup> appear to be disordered, delocalized and partially occupied. The electron density maps (Fig. 8b) suggest these water molecules occupy broad potential energy wells aligned along the groove. At the CpG step on the floor of the minor



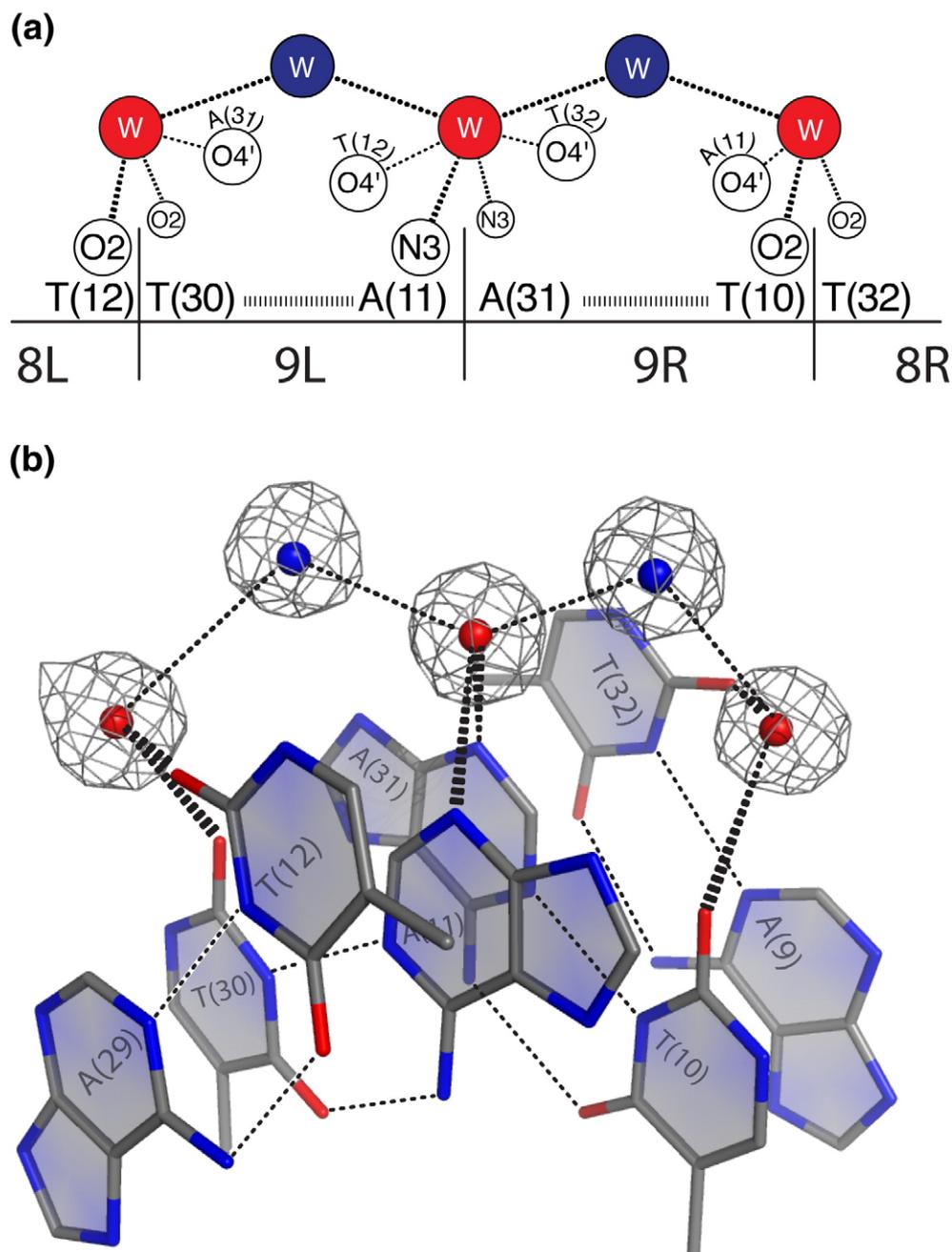
**Fig. 6.** Electrostatic surface potential of P22R NTD. Graphical representation of the ESP calculated at the solvent-accessible surface of the P22R NTD alone. The black circles indicate the phosphate pockets. P22R NTD is shown as a surface colored red ( $-3 \text{ kT/e}^-$ ) to blue ( $+3 \text{ kT/e}^-$ ). The DNA backbone is shown in tubes with phosphate oxygen atoms as red spheres. The DNA bases are shown as a cartoon.

groove in the central region, a continuous tube of electron density links sites within the primary and secondary layers of hydration. For some water molecules the electron density is weak and/or bimodal. The distances between water molecules in the model show greater variation ( $2.93 \pm 0.47$ ) than in DNA<sup>9T</sup>.

In DNA<sup>9C</sup> (Fig. 8) the central water molecule of the primary layer bridges the N3 atoms of G(11) and G(31) and also forms hydrogen bonds with two O4' atoms and with two secondary-layer water molecules. This water molecule can interact with multiple arrays of hydrogen-bonding partners in closely spaced sites. One site allows hydrogen bonding with the N2 atom of G(11), while another site allows hydrogen bonding with N2 of G(31). The

tube-like electron density observed in the minor groove of DNA<sup>9C</sup> (Fig. 8b) is consistent with partial water occupancy at a series of closely spaced sites, contrasting with the well-defined spheres of electron density associated with the DNA<sup>9T</sup> (compare Figs. 7b and 8b).

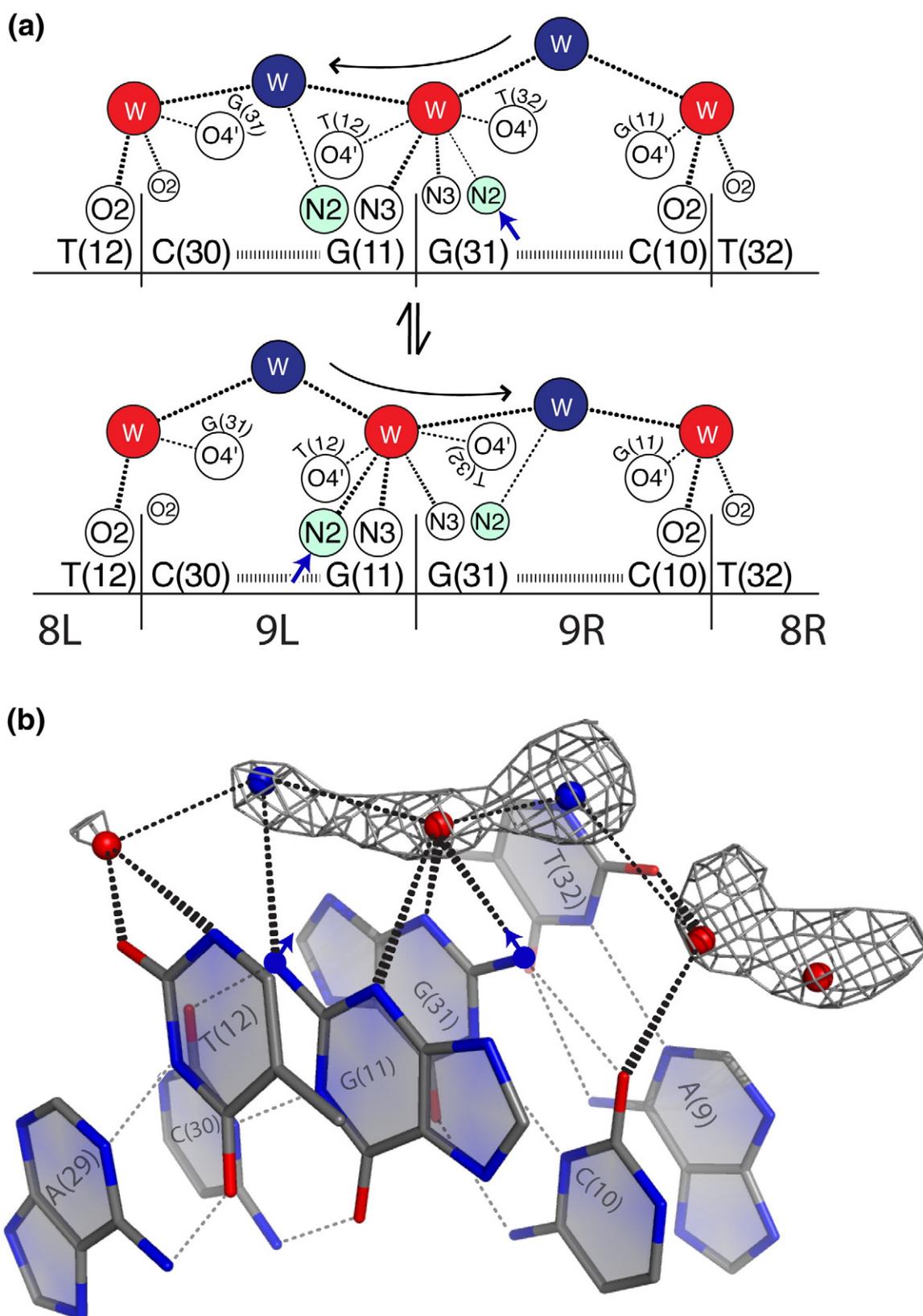
When the central primary water molecule forms a hydrogen bond with the N2 of G(31), a secondary water molecule forms a hydrogen bond with the N2 of G(11). Conversely, when the primary water molecule shifts to form a hydrogen bond with the N2 of G(11) a secondary water molecule shifts to form a hydrogen bond with the N2 of G(31). These translations in the central region in turn affect the outlying primary water molecules.



**Fig. 7.** A TpA step: induced B' state with a zigzag spine of hydration. A spine of hydration is observed in the minor groove of the non-contacted region of the P22R NTD-DNA<sup>9T</sup> complex. (a) A schematic diagram illustrating the hydrogen bonds, which are shown as dashed lines. The hydrogen-bond acceptors of the DNA minor groove are indicated by white spheres containing the appropriate atom labels. The water molecules are colored in accordance with those of Fig. 3. (b) Electron density defining the spine of hydration ( $2F_o - F_c$  map contoured at  $1.1\sigma$ ).

Several lines of evidence indicate that the tube-like electron density of the hydration layer in the central region of DNA<sup>9C</sup> (Fig. 8b) is a characteristic of the CpG step in the DNA<sup>9C</sup> complex and does not result from issues related to data quality or crystallization conditions. First, the upper layers of hydration in the non-contacted region of P22R NTD-DNA<sup>9C</sup> complex are as well defined as those in the P22R NTD-DNA<sup>9T</sup> complex. Second, the primary and secondary layers of the spine of hydration in the

intrinsic A-tract present in the contacted region (positions 1L-4L and 1R-4R) of the DNA<sup>9C</sup> and DNA<sup>9T</sup> complex are equally well defined. Finally, the P22R NTD-DNA<sup>9C</sup> complex has been solved in a different space group ( $P4_3$ ; data not shown) at 2.0 Å resolution, and the primary and secondary layer of the spine of hydration in the non-contacted region there shows the identical distortions of electron density, in particular a tube of density linking the three central water molecules.



**Fig. 8.** A CpG step: induced B' state with a delocalized zigzag spine of hydration. The spine of hydration in the minor groove of the non-contacted region in the P22R NTD–DNA<sup>9C</sup> complex is represented in analogy with that of Fig. 7. (a) A schematic diagram representing the conversion between two hydration states. The hydrogen-bond donors (2-amino groups of G) distinguish DNA<sup>9C</sup> from DNA<sup>9T</sup> and are shaded green. In each state, the hydrogen-bond donor that participates in hydrogen-bonding state is indicated by a blue arrow. The curved arrows indicate transitions between the alternative hydration states. (b) Observed electron density showing the average of the two hydration states ( $2F_o - F_c$  map contoured at  $1.1\sigma$ ).

## Counterions

Mg<sup>2+</sup> ions can be inferred from coordination geometry. Careful inspection of the electron density maps does not reveal evidence for localization of Mg<sup>2+</sup> ions. Monovalent cations (Na<sup>+</sup> and K<sup>+</sup>) are indistinguishable from water molecules and are generally observable only by substitution with Rb<sup>+</sup> or Tl<sup>+</sup>.<sup>24,25,31,47–49</sup> Therefore, the P22R NTD–DNA<sup>9T</sup> complex was crystallized from solutions containing either Tl<sup>+</sup> or Rb<sup>+</sup>. The Tl<sup>+</sup> and Rb<sup>+</sup> results are self-consistent except that the Tl<sup>+</sup> signal is significantly stronger than that of Rb<sup>+</sup>. Isomorphous difference and anomalous methods from the Tl<sup>+</sup> data were used to determine positions of localized monovalent cations. Anomalous difference maps (from P22R NTD–DNA<sup>9T</sup>/Tl<sup>+</sup> data), synthesized using  $|F_+ - F_-|$  as Fourier coefficients, generally confirm the isomorphous difference results, although the signal is somewhat weaker. Isomorphous difference maps were calculated using the diffraction data from the P22R NTD–DNA<sup>9T</sup>/Na<sup>+</sup> derivative<sup>2</sup> and a P22R NTD–DNA<sup>9T</sup>/Tl<sup>+</sup> derivative.

The results indicate unambiguous localization of monovalent cations (Fig. 9). The Tl<sup>+</sup>–Na<sup>+</sup> isomorphous difference peaks define the positions of at least four localized monovalent cations per complex. Three of these peaks are located in the major groove, adjacent to guanine bases. One of these ions is adjacent to G(28) at position 7L, and two are adjacent to the terminal base G(40). The Tl<sup>+</sup> at G(28) is displaced by the amino group of a lysine at position G(8) [the pseudo-symmetry mate G(28); Fig. 9a and 9b]. The lysine is contributed by a symmetry-related protein molecule. The two cation sites at the terminal base G(40) likely represent a single cation that partially occupies two positions.

A clear isomorphous difference peak (Fig. 9e) is located in the minor groove of one of the intrinsic A-tracts, which spans positions 1R–4R. This peak is observed in the R half site, which is more highly ordered than the L half site, giving higher-quality maps. This peak indicates a partially occupied monovalent cation is coordinated by the A(N3) and T(O2) functional groups on the floor of the groove and by the O4' atoms of flanking nucleotides. This mode of minor groove coordination of monovalent cations has been observed previously.<sup>25,32,33,48,50,51</sup> Weaker peaks at adjacent sites on the floor of the A-tract suggest additional cations. In general, it can be seen that positions of monovalent cations correspond to solvent sites (water molecules) indicated in the  $2F_o - F_c$  maps of P22R NTD–DNA<sup>9T</sup>. Two of the major groove anomalous peaks are near the junction between two symmetry-related DNA duplexes (Fig. 9c and 9d).

Monovalent cations are known to associate with the minor groove of A-tracts, and the ESP calculations suggest that monovalent cations might be found in the minor groove at the center of the binding site. Therefore, we looked for monovalent cation binding sites in the non-contacted region of the P22R NTD–DNA<sup>9C</sup> complex. No evidence of

monovalent cation localization was evident in the isomorphous difference or anomalous maps.

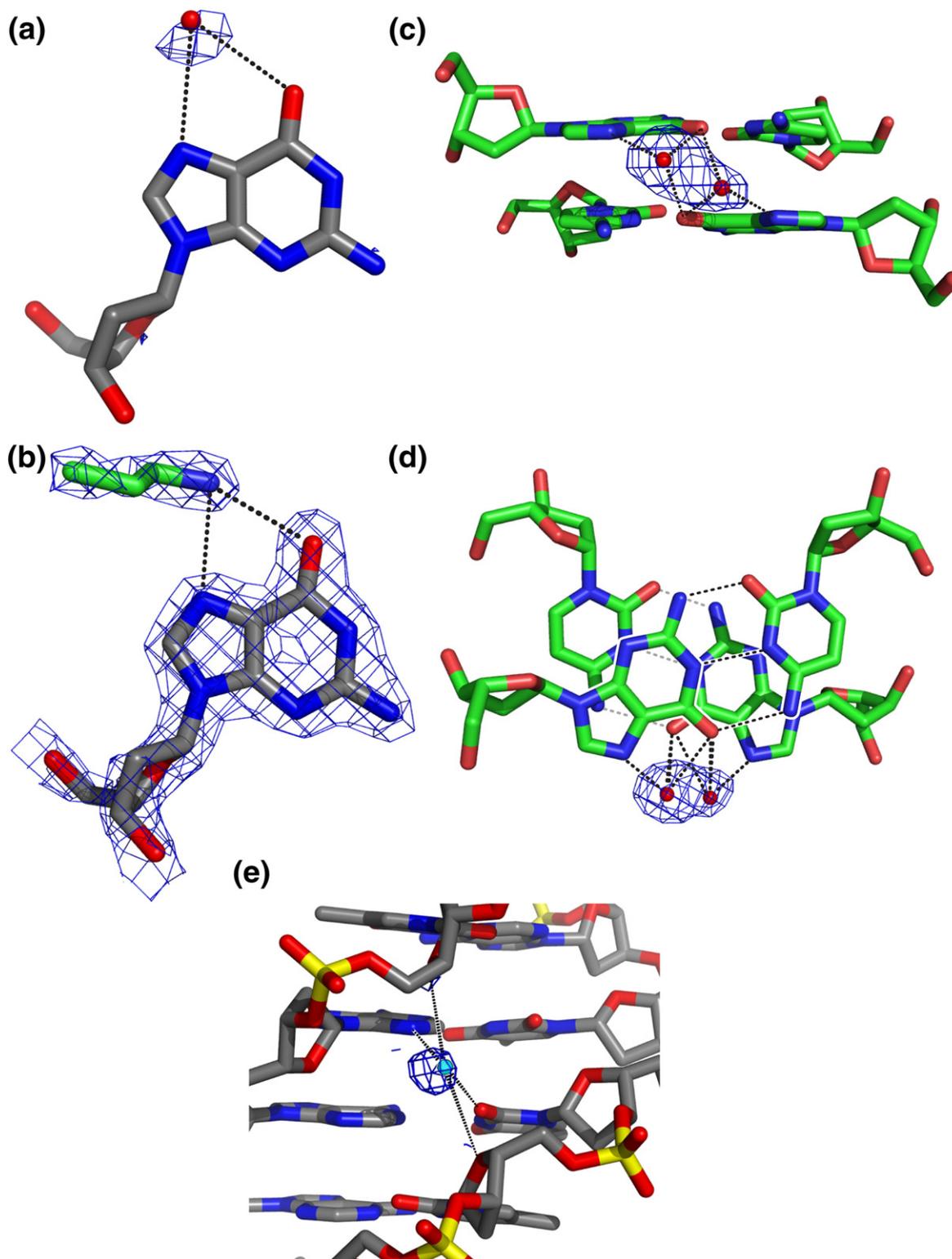
## Discussion

Indirect readout is an important and common mechanism for DNA recognition. Indirect readout is the primary method of DNA sequence detection used by protein assemblies such as nucleosomes of eukarya<sup>14</sup> and archaea,<sup>52,53</sup> and IHF and HU complexes of bacteria.<sup>54,55</sup> Indirect and direct readout of DNA are observed in bacteria and bacterial viruses,<sup>56–59</sup> eukaryotes,<sup>60–65</sup> as well as the viruses of eukaryotic cells.<sup>66,67</sup>

### A structural and thermodynamic model of indirect readout

We have presented a mechanistic model of indirect readout based on the high-resolution structure<sup>2</sup> of P22R NTD in complex with operator DNA<sup>9T</sup> and over a decade's accumulation of biochemical data.<sup>10–13,58,68,69</sup> This model suggests that regardless of its sequence, the non-contacted region of the P22 operator is forced into the B' state by the repressor. The sequence is "read" by the ease with which it assumes the B' state. This recognition model makes testable predictions about the structures and stabilities of P22R complexes with a variety of DNA sequences. The model predicts conservation of the global architecture of the protein–DNA complex, conservation of the conformation of the protein and the DNA, and conservation of the protein–DNA contacts as the sequence of the non-contacted region is varied. The model predicts the central CpG step of DNA<sup>9C</sup> assumes the B' state, even though C–G base pairs have not been observed previously in the B' state. Finally, the model predicts that the affinity of DNA for P22R NTD is modulated by the resistance of the non-contacted sequence to the transition to the B' state. The stability of the B' state varies with sequence (ApA, ApT>TpA>CpG, CpC, GpC, GpG). The affinity should be greatest for A-tracts and lowest for G/C tracts. Indeed sequences that are most resistant to the B' state<sup>3</sup> have lowest affinity for P22R.<sup>12,13</sup> The structure of the P22R NTD–DNA<sup>9C</sup> complex supports the predictions of this model.

The DNA binding surface of P22R NTD is electrostatically complementary to the phosphate group disposition of B'–DNA, but not of B–DNA. ESP calculations reveal pockets of positive potential near the dimer interface (Fig. 6) that juxtapose with B'–DNA phosphates. Two negatively charged amino acids, E44 and E48, are in proximity with the positive pockets of the repressor. These residues may "guide" the DNA phosphate backbone along the surface of P22R, destabilizing canonical B-form DNA relative to the B' state. Consistent with this idea, an E44A mutation, which removes the negative charge at this position, eliminates P22R's ability to discriminate between DNA<sup>9C</sup> and DNA<sup>9T</sup>. An



**Fig. 9.** Positions of localized counter ions. (a) A  $\text{Ti}^+$  ion is coordinated by O6 and N7 of G(28) of the L-half of the operator (7L). The position of the  $\text{Ti}^+$  ion is indicated by an anomalous difference peak contoured at  $2.5\sigma$ . (b) A lysine at position G(8) on the R-half (7R) of the operator has displaced a  $\text{Ti}^+$  ion. G(8) is the pseudo-symmetry mate of G(28). The positions of the atoms are indicated by a sum map, contoured at  $2.0\sigma$ . (c)  $\text{Ti}^+$  ions in the major groove at the junction between two DNA molecules. The positions of these ions are indicated by anomalous peaks, contoured at  $10\sigma$ . (d) Axial view of the ions and maps shown in (c). (e) A  $\text{Ti}^+$  ion in the minor groove of an A-tract of the contacted region. The position of this ion is indicated by an  $F_{\text{Ti}^+} - F_{\text{Na}^+}$  isomorphous difference peak, contoured at  $2.5\sigma$ .

E44D mutant, which does not remove the negative charge at this position, retains discrimination between these two sequences (L. Harris and G.B. K., unpublished results).

### The B' state: role of minor groove hydration

The minor groove of A-tract DNA is a unique hydration environment, which has no obvious parallels elsewhere. A curved sheet, one molecule thick (Fig. 3a), contains discrete, optimally spaced water molecules, which are translationally and rotationally restrained (Fig. 3). The rotational restraints arise from an absence of hydrogen-bond donors, a hallmark of the minor groove of A-tract DNA.

The water molecules in the A-tract minor groove are linked and interdependent.<sup>14,24,26,70,71</sup> The hydration assembly is a "house of cards" in that it requires correct positions, rotations and occupancies of each water molecule. Perturbations of position, rotation and/or occupancy anywhere within the foundation of the assembly destabilize B'-DNA.

The barriers to translation of water molecules within the hydration assembly of A-tract DNA arise from the positioning of multiple hydrogen-bond acceptors of the DNA. The barriers to rotation arise from a donor-acceptor imbalance of the DNA. A-tract DNA lacks hydrogen-bond donors in the minor groove. To form optimum hydrogen bonds, water molecules in the primary layer must direct their protons down toward the floor of the groove (Fig. 3b). The water molecules in the primary layer exclusively donate protons to acceptors on the floor and walls of the groove. Each proton is involved in two bifurcated hydrogen bonds (i.e., two "three-centered hydrogen bonds"; for a discussion of bifurcated hydrogen bonds, see Taylor *et al.*<sup>72</sup>) in which each proton interacts with two hydrogen-bond acceptor atoms. The water molecules of the primary layer are polarized, with parallel dipole moments. The polarization extends to other layers; water molecules in the second layer direct their protons down toward the primary layer.

The foundation of the hydration assembly is a zigzag spine of hydration,<sup>22,23</sup> which forms a base for additional layers.<sup>24-26</sup> Each primary layer water molecule forms a bridge between DNA strands (Fig. 3b). These bridges involve both sugars and bases. In A-tracts, which contain only ApA and ApT steps, the bridges link (at an ApA step) the N3 (A) of the one strand with the O2 (T) of the other strand, or (at an ApT step) two O2 (T) atoms. As described below, when the sequence of B'-DNA diverges from that of canonical A-tracts, the hydrogen-bonding interactions and the rotational and positional restraints of the primary hydration layer are altered

### The B' state: effect of DNA sequence

The B' state forms spontaneously in A-tract DNA (Fig. 3). Non-A-tract sequences can be induced or forced into the B' state. For example, as illustrated in

Fig. 7, the zigzag hydration assembly is observed in the non-contacted minor groove of the P22R NTD-DNA<sup>9T</sup> complex.<sup>2</sup> Although this sequence is not an A-tract, the protein has induced the DNA conformation and hydration corresponding to the B' state. The primary water molecule at the central TpA step (positions 9L and 9R) is coordinated by two N3 (A) atoms, in analogy to the water coordination at an A-tract ApT step that involves two O2 (T) atoms.

The TpA→CpG substitution at positions 9L and 9R (converting DNA<sup>9T</sup> to DNA<sup>9C</sup>) does not prevent the formation of the B' state in the P22R-bound DNA operator. In the complex, the DNA<sup>9C</sup> non-contacted minor groove is narrow, with a spine-like hydration assembly. However, the TpA→CpG substitution introduces two amino groups, which are hydrogen-bond donors, to the floor of the minor groove. Because the primary water molecules interact with both hydrogen-bond donors and acceptors, the positional and rotational restraints on these water molecules are released. In the DNA<sup>9C</sup> complex, the central primary water molecule shifts between the two 2-amino groups, forming hydrogen bonds with one N2, or the other N2, but not both (Fig. 8a). The delocalization of the primary water molecules appears to disturb the entire hydration assembly, with water molecules in the secondary layer shifting in concert with the those in the primary layer.

These findings suggest that the water molecules in the non-contacted minor groove in P22R NTD-DNA<sup>9C</sup> complex molecules are more mobile, and less stable, than analogous solvent molecules in the P22R NTD-DNA<sup>9T</sup> complex. A less stable hydration assembly after the TpA→CpG mutation is consistent with the 10-fold lower affinity of P22R. DNA sequence modulates the stability and structure of the B'-specific minor groove hydration assembly that in turn regulates affinity for P22R.

### The B' state: solution experiments

The structural results suggest that water molecules in the minor groove of the P22R NTD-DNA<sup>9C</sup> complex are more mobile than those in the P22R NTD-DNA<sup>9T</sup> complex. This prediction is testable by <sup>•</sup>OH footprinting (Fig. 10).

<sup>•</sup>OH is a small diffusible footprinting agent that cleaves DNA by attacking deoxyribose sugars from within the minor groove (Fig. 10a).<sup>74</sup> The rate of strand scission of a given sugar is thought to depend on the steric accessibility. In P22R-DNA complexes, diffusion through the tunnel is the only path for <sup>•</sup>OH radical to enter the minor groove (Fig. 10c). Steric effects do not change with DNA sequence; the conformations are invariant. If steric effects dominate, accessibility of <sup>•</sup>OH radical into the DNA<sup>9C</sup> and DNA<sup>9T</sup> tunnels would not vary.

However, <sup>•</sup>OH cleaves within the DNA<sup>9C</sup> tunnel around fivefold more rapidly than within the DNA<sup>9T</sup> tunnel (Fig. 10b).<sup>10</sup> The simplest explanation for the rapid cleavage within the DNA<sup>9C</sup> tunnel is that the solvent milieu is more mobile, as illustrated in Fig. 8a, and that rates of diffusion are greater in

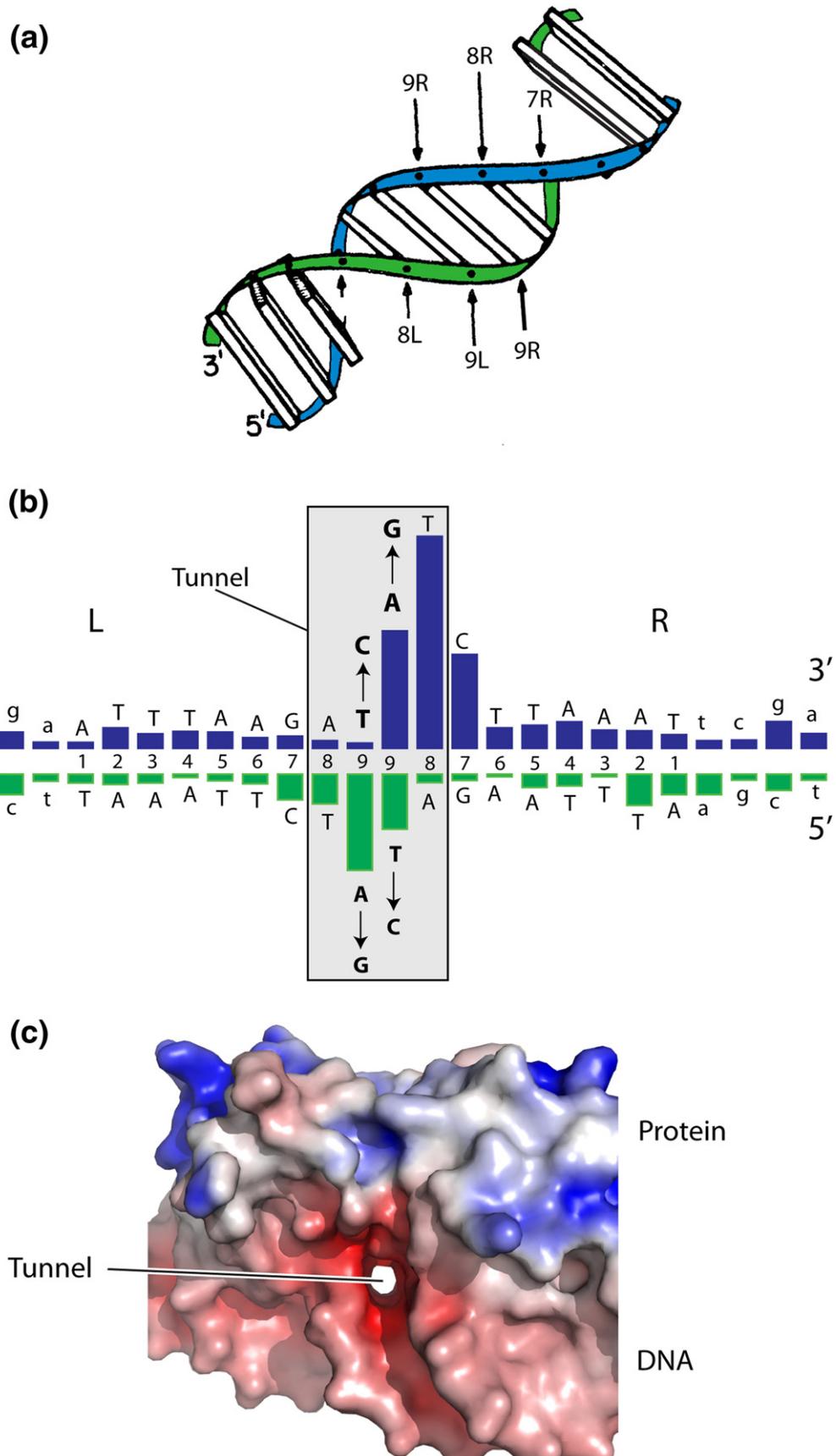


Fig. 10 (legend on next page)

the DNA<sup>9C</sup> tunnel than in the DNA<sup>9T</sup> tunnel. Therefore, the cleavage data supports the structure-based model of greater water mobility in the non-contacted region of DNA<sup>9C</sup> compared to DNA<sup>9T</sup>. C·G base pairs increase minor groove water mobility.

### Cation localization

As previously observed for DNA alone,<sup>30,48</sup> the complex with P22R NTD reveals monovalent cations in the major groove in association with C·G base pairs and in the minor groove in association with A·T base pairs of A-tracts. ESP calculations reveal that P22R binding increases the negative potential of the minor groove in the non-contacted region. However, no evidence of cation localization is observed in the P22R NTD–DNA<sup>9T</sup>/Ti<sup>+</sup> or the P22R NTD–DNA<sup>9C</sup>/Rb<sup>+</sup> complexes presented here. We believe that high levels of NH<sub>4</sub><sup>+</sup>, remaining from the protein purification, may have displaced the anomalous scatters.

### Conclusion

For P22R NTD to bind to the DNA operator, with DNA phosphate groups located in cationic pockets in the protein, the non-contacted minor groove must be narrow in the center of the operator. The narrow minor groove is characteristic of the B' state. The narrowing of the minor groove (and the increase in proximity of phosphate groups) is thermodynamically linked to the formation of a spine of hydration. The B→B' transition is sequence dependent. In free DNA, A-tracts have high propensities to conform to the B' state.<sup>31,34,75</sup> This propensity is partially due to the proper positioning of chelating O2 atoms of thymine to coordinate water molecules and cations.

We have shown that the binding of P22R NTD induces a narrow minor groove as well as a spine of hydration in non-A-tract DNA, despite the presence of destabilizing TpA or even CpG steps. The CpG step results in a disordered spine, particularly near the floor of the minor groove. The change in the spine of hydration is the only observable difference between the X-ray crystal structures of the DNA<sup>9C</sup> and the DNA<sup>9T</sup> complexes. We propose that the reduced stability of the spine results in a higher energy cost for the narrow minor groove in the DNA<sup>9C</sup> operator. The higher energy cost for a narrow minor groove results in a higher energy state of the complex. These findings are in agreement with the solution studies that show the P22 c2 repressor binds a DNA operator with a CpG at the

central non-contacted step with lower affinity than that with a TpA at the central step.

## Materials and Methods

### Preparation of protein and DNA

P22R NTD was isolated from *Escherichia coli* strain XA90<sup>76</sup> bearing a plasmid that directs overproduction of the amino acids 1–68 of the bacteriophage P22 c2 repressor. The construction of the P22R NTD overexpressing plasmid and general method for purification of P22R NTD were described previously.<sup>2</sup> For preparation of P22R NTD co-crystallized with DNA in the presence of Ti<sup>+</sup>, the final S-100 column was developed with a buffer containing 50 mM Tris–acetate (pH 7.5) and 100 mM thallium acetate. The eluted protein was precipitated with ammonium sulfate. The protein-containing pellet was dissolved in and subsequently dialyzed against the same buffer supplemented with 10% glycerol. A similar procedure was used for preparing proteins used in the Rb<sup>+</sup> crystal, except the S-100 column and the suspension and dialysis buffer consisted of 50 mM Tris–HCl (pH 7.5), 100 mM RbCl, and 10% glycerol. Following dialysis, purified protein was flash-frozen in liquid nitrogen and stored at –70 °C.

DNA oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA) as ammonium salts from reverse HPLC purification. To form double strands, equimolar amounts of the complementary strands were mixed in water, heated to 85 °C for 60 s and slow-cooled to room temperature over 4 h to allow for annealing of the two strands.

### Crystallization

Crystals were grown at 4 °C in hanging drops of equimolar P22R NTD–duplex DNA at 0.42 mM and equilibrated by vapor diffusion. Crystals grew to a size of approximately 0.5 mm × 0.3 mm × 0.3 mm within 1 day. Crystals were collected in nylon loops and flash-frozen in liquid nitrogen.

#### P22R NTD with DNA<sup>9C</sup>

The initial crystallization solution contained duplex d(5'-TATTTAAGACGTCTTAAATG-3')–d(5'-CATTTAA-GACGTCTTAAATA-3'), 45 mM Tris–HCl (pH 7.8), 20 mM NaCl, 2.0 mM glycerol, 14% polyethylene glycol (PEG) 400, 4.5 mM LiCl, 2.3 mM MgCl<sub>2</sub> and 0.9% methyl-2,4-pentanediol (MPD) in a volume of 5.3 μl. The reservoir contained 100 mM Tris–HCl (pH 7.8), 30% PEG 400, 10 mM LiCl, 5 mM MgCl<sub>2</sub> and 2% MPD.

#### P22R NTD with DNA/Ti<sup>+</sup>

The initial crystallization solution contained DNA<sup>9T</sup> duplex d(5'-CATTTAAGATATCTTAAATG-3')–d(5'-

**Fig. 10.** C·G base pairs increase mobility within the tunnel. (a) A schematic diagram illustrating the known 3' offset of cleavage of two strands of DNA by a small diffusible agent that cuts within the minor groove. This panel is adapted from Dervan.<sup>73</sup> (b) Hydroxyl radical footprinting differences between DNA<sup>9T</sup> and DNA<sup>9C</sup> in complexes with P22R NTD. DNA<sup>9C</sup> shows elevated cleavage within the tunnel region. The rates of solvent transit within the tunnel are greater for DNA<sup>9C</sup> than for DNA<sup>9T</sup>. This panel is adapted from Wu *et al.*<sup>13</sup> (c) The tunnel of the P22R NTD–DNA<sup>9C</sup> complex, showing the ESP. This view is looking along the minor groove, through the tunnel. The surface is colored red (–10 kT/e<sup>–</sup>) to blue (+3 kT/e<sup>–</sup>).

CATTTAAGATATCTTAAATG-3'), 20 mM thallium acetate, 45 mM Tris-acetate (7.8), 2.0% glycerol, 14% PEG 400, 2.3 mM magnesium acetate and 0.9% MPD in a volume of 5.7  $\mu$ l. The crystallization solution was equilibrated against a reservoir of 100 mM Tris-acetate (pH 7.8), 25% PEG 400, 5 mM magnesium acetate and 2% MPD. Similar protocols were used in attempt to co-crystallize P22R NTD with duplex DNA<sup>9C</sup>; however, for this DNA, TI<sup>+</sup> cations were not observed in either anomalous or isomorphous difference maps.

#### P22R NTD with DNA<sup>9C</sup>/Rb<sup>+</sup>

The initial crystallization solution contained duplex d (5'-CATTTAAGACGCTTAAATG-3')-d(5'-CATTTAAGACGCTTAAATG-3'), 45 mM Tris-HCl (pH 7.8), 20 mM rubidium chloride, 2.0% glycerol, 14% PEG 400, 2.3 mM magnesium chloride and 0.9% MPD in a volume of 5.3  $\mu$ l. The reservoir contained 100 mM Tris-HCl (pH 7.8), 25% PEG 400, 5 mM magnesium chloride and 2% MPD. Similar protocols were used in an attempt to co-crystallize P22R NTD with duplex DNA<sup>9T</sup>; however, for this DNA, Rb<sup>+</sup> cations were not observed in either anomalous or isomorphous difference maps.

#### Data collection

##### P22R NTD-DNA<sup>9C</sup>

Diffraction data were collected on beamline X26-C at The National Synchrotron Light Source, Brookhaven National Laboratory, with an ADSC Quantum 4 CCD detector using radiation of 0.99997 Å wavelength, with the crystal at 110 K. A total of 232,653 reflections were indexed

and integrated with DENZO<sup>77</sup> and reduced to 32,768 unique reflections in space group  $P2_12_12_1$  in the resolution range of 35–1.60 Å. Data were scaled with the program SCALEPACK.<sup>77</sup>

##### P22R NTD-DNA<sup>9T</sup>/TI<sup>+</sup>

Diffraction data were collected on SER-CAT beamline 22 BM at the Advanced Photon Source, Argonne National Laboratory, with a MAR225 CCD detector using radiation of 0.97784 Å wavelength, with the crystal at 110 K. A total of 452,190 reflections were indexed, integrated and reduced to 30,146 unique reflections in space group  $P4_3$  in the resolution range of 30–1.9 Å.

##### P22R NTD-DNA<sup>9C</sup>/Rb<sup>+</sup>

Diffraction data were collected on SER-CAT beamline 22 BM at the Advanced Photon Source, Argonne National Laboratory, with a MAR225 CCD detector using radiation of 0.81398 Å, with the crystal at 110 K. A total of 347,031 reflections were indexed, integrated and reduced to 22,529 unique reflections in space group  $P4_3$  in the resolution range of 41–2.1 Å. Data were scaled with the SCALEPACK with Bijvoets pairs unmerged.

#### Refinement

Initial phases were determined by molecular replacement with the program Molrep of the CCP4 suite<sup>78</sup> using data from 15 to 3 Å. The P22R NTD-DNA<sup>9T</sup> complex with water molecules removed was used a search model. Model building was performed with the program

**Table 1.** Data collection and refinement statistics

	P22R NTD-DNA <sup>9C</sup>	P22R NTD-DNA <sup>9T</sup> /TI <sup>+</sup>	P22R NTD-DNA <sup>9C</sup> /Rb <sup>+</sup>
X-ray source	X26-C NSLS	SER-CAT 22-BM APS	SER-CAT 22-BM APS
Wavelength (Å)	0.99997	0.97784	0.81398
Detector	ADSC Quantum 4 CCD	MAR CCD 225	MAR CCD 225
Resolution range (Å)	19.90–1.67	30.00–1.90	41.00–2.10
Mosaic spread (°)	~0.7	~0.5	~0.7
Space group	$P2_12_12_1$	$P4_3$	$P4_3$
Unit-cell parameters (Å)			
<i>a</i>	40.602	64.249	64.007
<i>b</i>	54.319	64.249	64.007
<i>c</i>	114.112	100.82	101.623
Total reflections	232,653	452,190	347,031
Unique reflections	32,768	30,146	22,529
Multiplicity <sup>a</sup>	5.8 (4.7)	13.4 (11.3)	12.5 (10.4)
Completeness (%)	97.3 (91.7)	98.5 (97.7)	99.3 (99.6)
Highest-resolution shell (Å)	1.67–1.71	1.95–1.90	2.15–2.10
<i>I</i> / $\sigma$ ( <i>I</i> )	47.1 (2.12)	44.3 (9.29)	33.2 (4.5)
<i>R</i> <sub>sym</sub> (%)	6.9 (65.9)	7.3 (48.7)	14.2 (63.6)
Reflections (working/test set)	27,753/1501	30,146/1591	22,529/1214
No. of non-H protein atoms	2 × 514	2 × 514	2 × 514
No. of non-H DNA atoms	814	814	814
No. of water molecules	330	304	296
No. of cations <sup>b</sup>		4	2
<i>R</i> <sub>crystal</sub> (%)	19.78	18.53	18.00
<i>R</i> <sub>free</sub> (%)	22.45	20.83	22.50
Average <i>B</i> -factor	24.78	25.06	26.55
RMSD from ideal			
Bond lengths (Å)	0.01	0.01	0.01
Bond angles (°)	1.6	1.7	2.2

<sup>a</sup> The number in parentheses indicates the value for the highest-resolution shell.

<sup>b</sup> The number of cations included in the final model/number of cations indicated by anomalous maps and/or isomorphous difference maps.

COOT.<sup>79</sup> Water molecules were added iteratively to the model initially with ARP/wARP<sup>80</sup> and by inspection of  $|2F_o - F_c|$  and  $|F_o - F_c|$  maps in later stages, followed by cycles of refinement in Refmac5.<sup>81</sup> Data collection and refinement statistics are given in Table 1.

#### P22R NTD-DNA<sup>9C</sup>

Following rigid-body refinement, difference density was observed in the  $|F_o - F_c|$  maps at positions 9L and 9R of the operator corresponding to T·A→C·G mutations at these positions. After correction of the model, addition of water and final rounds of TLS refinement in Refmac, the final model converged at 1.67 Å resolution with an *R*-factor of 19.78% and an *R*<sub>free</sub> of 22.58%.

#### P22R NTD-DNA<sup>9T</sup>/Ti<sup>+</sup>

The final *R*-factor is 18.4% and the *R*<sub>free</sub> is 20.8%.

#### P22R NTD-DNA<sup>9C</sup>/Rb<sup>+</sup>

The final *R*-factor is 18.3% and the *R*<sub>free</sub> is 22.4%.

#### Anomalous ( $F_+$ - $F_-$ ) and isomorphous ( $F_{Ti^+}$ - $F_{Na^+}$ ) difference maps

Anomalous ( $F_+$  -  $F_-$ ) maps (Fig. 9a, c and d) were calculated for P22R NTD-DNA<sup>9T</sup>/Ti<sup>+</sup> and P22R NTD-DNA<sup>9C</sup>/Rb<sup>+</sup> complexes using FFT from CCP4.<sup>78</sup> Phases for the anomalous maps were calculated from models with solvent molecules removed, using Sfall of CCP4. Weak anomalous peaks corresponding to phosphorous atoms of the DNA backbone indicate that the maps were calculated accurately.

Isomorphous difference ( $|F_{Ti^+} - F_{Na^+}|$ ) maps (Fig. 9e) were calculated using the native data from the P22R NTD-DNA<sup>9T</sup>/Na<sup>+</sup> complex.<sup>2</sup> P22R NTD-DNA<sup>9T</sup>/Ti<sup>+</sup> data were scaled to the P22R NTD-DNA<sup>9T</sup>/Na<sup>+</sup> data using Scaleit from CCP4. Phases were calculated from the P22R NTD-DNA<sup>9T</sup>/Ti<sup>+</sup> model with all solvent sites treated as water molecules. Clear peaks indicated the positions of localized Ti<sup>+</sup> ions. Isomorphous difference ( $|F_{Rb^+} - F_{Na^+}|$ ) maps were calculated with the P22R NTD-DNA<sup>9C</sup>/Rb<sup>+</sup> data. However, presumably due to the weaker scattering of Rb<sup>+</sup> compared to Ti<sup>+</sup>, the signal was not significantly greater than noise. The observed isomorphous difference peaks are located in the positions that overlap with either those identified from anomalous scattering data or are predicted from a previous structures of DNA oligomers crystallized in the presence of Ti<sup>+</sup>.<sup>47,48,50</sup> One of the Ti<sup>+</sup> sites observed in the isomorphous difference maps is located in the minor groove linking the bases of positions 2R and 3R and is coordinated by the O2 of T(24) and the N3 A(18) (Fig. 9e).

In comparison with previous work,<sup>47,48</sup> the statistics indicate that the Na<sup>+</sup> and Ti<sup>+</sup> complexes here are highly isomorphous. The data sets merged with an *R*<sub>iso</sub><sup>82</sup> of 28.5% averaged over all shells. Differences in unit cell axis are  $\Delta a = 0.2\%$ ,  $\Delta b = 0.2\%$ , and  $\Delta c = 0.8\%$ . The RMSD for all backbone atoms of the DNA and protein after refinement is 0.41 Å.

#### Electrostatic calculations

ESP calculations were performed on the P22R NTD alone (Fig. 6), DNA<sup>9T</sup> alone, DNA<sup>9C</sup> alone, P22R NTD-

DNA<sup>9T</sup> and P22R NTD-DNA<sup>9C</sup> by solving the nonlinear Poisson-Boltzmann equation using the Adaptive Poisson Boltzmann Solver (APBS)<sup>83</sup> with the Pymol<sup>84</sup> interface. Calculations were run at 150 mM NaCl with a probe size of 1.4 Å. The van der Waals radii and partial charges of the atoms of the protein and the DNA were estimated using Amber as implemented in PDB2PQR.<sup>85</sup>

#### PDB accession numbers

Coordinates and structure factors have been deposited into the PDB with accession numbers 3JXB (P22R NTD-DNA<sup>9C</sup>), 3JXC (P22R NTD-DNA<sup>9T</sup>/Ti<sup>+</sup>) and 3JXD (P22R NTD-DNA<sup>9C</sup>/Rb<sup>+</sup>).

#### Acknowledgements

The authors thank Drs. Anton Petrov, Stephen Harvey, Nicholas Hud, Roger Wartell and Victor Zhurkin for helpful discussions.

#### References

- Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA*, **73**, 804–808.
- Watkins, D., Hsiao, C., Woods, K. K., Koudelka, G. B. & Williams, L. D. (2008). P22 c2 repressor-operator complex: mechanisms of direct and indirect readout. *Biochemistry*, **47**, 2325–2338.
- Hud, N. V. & Plavec, J. (2003). A unified model for the origin of DNA sequence-directed curvature. *Biopolymers*, **69**, 144–158.
- Rich, A. (1993). DNA comes in many forms. *Gene*, **135**, 99–109.
- Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M. & Zhurkin, V. B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl Acad. Sci. USA*, **95**, 11163–11168.
- McFail-Isom, L., Sines, C. C. & Williams, L. D. (1999). DNA structure: cations in charge? *Curr. Opin. Struct. Biol.* **9**, 298–304.
- Koudelka, G. B. & Carlson, P. (1992). DNA twisting and the effects of non-contacted bases on affinity of 434 operator for 434 repressor. *Nature*, **355**, 89–91.
- Zhang, Y., Xi, Z., Hegde, R. S., Shakked, Z. & Crothers, D. M. (2004). Predicting indirect readout effects in protein-DNA interactions. *Proc. Natl Acad. Sci. USA*, **101**, 8337–8341.
- Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S. & Honig, B. (2009). The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
- Wu, L. & 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.
- Hilchey, S. P., Wu, L. & 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.
- Hilchey, S. P. & 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.
13. Wu, L., Vertino, A. & Koudelka, G. B. (1992). Non-contacted bases affect the affinity of synthetic P22 operators for P22 repressor. *J. Biol. Chem.* **267**, 9134–9139.
  14. Segal, E. & Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* **19**, 65–71.
  15. Arnott, S. & Hukins, D. W. L. (1972). Optimized parameters for A-DNA and B-DNA. *Biochem. Biophys. Res. Commun.* **47**, 1504–1509.
  16. Burkhoff, A. M. & Tullius, T. D. (1987). The unusual conformation adopted by the adenine tracts in kinetoplast DNA. *Cell*, **48**, 935–943.
  17. Burkhoff, A. M. & Tullius, T. D. (1988). Structural details of an adenine tract that does not cause DNA to bend. *Nature*, **331**, 455–457.
  18. Price, M. A. & Tullius, T. D. (1993). How the structure of an adenine tract depends on sequence context: a new model for the structure of T<sub>n</sub>A<sub>n</sub> DNA sequences. *Biochemistry*, **32**, 127–136.
  19. Mack, D. R., Chiu, T. K. & 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.
  20. Tevis, D. S., Kumar, A., Stephens, C. E., Boykin, D. W. & Wilson, W. D. (2009). Large, sequence-dependent effects on DNA conformation by minor groove binding compounds. *Nucleic Acids Res.* **37**, 5550–5558.
  21. Alexeev, D. G., Lipanov, A. A. & Skuratovskii, I. Y. (1987). Poly(dA)-poly(dT) is a B-type double helix with a distinctively narrow minor groove. *Nature*, **325**, 821–823.
  22. Kopka, M. L., Fratini, A. V., Drew, H. R. & Dickerson, R. E. (1983). Ordered water structure around a B-DNA dodecamer. a quantitative study. *J. Mol. Biol.* **163**, 129–146.
  23. Drew, H. R. & Dickerson, R. E. (1981). Structure of a B-DNA dodecamer. III. Geometry of hydration. *J. Mol. Biol.* **151**, 535–556.
  24. Shui, X., Sines, C., McFail-Isom, L., VanDerveer, D. & Williams, L. D. (1998). Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry*, **37**, 16877–16887.
  25. Tereshko, V., Minasov, G. & Egli, M. (1999). A “hydrat-ion” spine in a B-DNA minor groove. *J. Am. Chem. Soc.* **121**, 3590–3595.
  26. Woods, K. K., Maehigashi, T., Howerton, S. B., Sines, C. C., Tannenbaum, S. & Williams, L. D. (2004). High-resolution structure of an extended A-tract: [d(CGCAAATTTGCG)]<sub>2</sub>. *J. Am. Chem. Soc.* **126**, 15330–15331.
  27. Marini, J. C., Levene, S. D., Crothers, D. M. & Englund, P. T. (1982). Bent helical structure in kinetoplast DNA. *Proc. Natl Acad. Sci. USA*, **79**, 7664–7668.
  28. Wu, H.-M. & Crothers, D. M. (1984). The locus of sequence-directed and protein-induced DNA bending. *Nature*, **308**, 509–513.
  29. Zinkel, S. S. & Crothers, D. M. (1987). DNA bend direction by phase sensitive detection. *Nature*, **328**, 178–181.
  30. Woods, K., McFail-Isom, L., Sines, C. C., Howerton, S. B., Stephens, R. K. & Williams, L. D. (2000). Monovalent cations sequester within the A-tract minor groove of [d(CGCGAATTCGCG)]<sub>2</sub>. *J. Am. Chem. Soc.* **122**, 1546–1547.
  31. Shui, X., McFail-Isom, L., Hu, G. G. & Williams, L. D. (1998). The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry*, **37**, 8341–8355.
  32. Hud, N. V., Sklenar, V. & 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.
  33. Cesare Marincola, F., Denisov, V. P. & Halle, B. (2004). Competitive Na(+) and Rb(+) binding in the minor groove of DNA. *J. Am. Chem. Soc.* **126**, 6739–6750.
  34. Wing, R., Drew, H., Takano, T., Broka, C., Takana, S., Itakura, K. & Dickerson, R. E. (1980). Crystal structure analysis of a complete turn of B-DNA. *Nature*, **287**, 755–758.
  35. Dickerson, R. E. & Drew, H. R. (1981). Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure. *J. Mol. Biol.* **149**, 761–786.
  36. Rhodes, D. (1979). Nucleosome cores reconstituted from poly (dA-dT) and the octamer of histones. *Nucleic Acids Res.* **6**, 1805–1816.
  37. Simpson, R. T. & Kunzler, P. (1979). Chromatin and core particles formed from the inner histones and synthetic polydeoxyribonucleotides of defined sequence. *Nucleic Acids Res.* **6**, 1387–1415.
  38. Herrera, J. E. & Chaires, J. B. (1989). A premelting conformational transition in poly(dA)-poly(dT) coupled to daunomycin binding. *Biochemistry*, **28**, 1993–2000.
  39. Chan, S. S., Breslauer, K. J., Austin, R. H. & Hogan, M. E. (1993). Thermodynamics and premelting conformational changes of phased (dA)<sub>5</sub> tracts. *Biochemistry*, **32**, 11776–11784.
  40. Augustyn, K. E., Wojtuszewski, K., Hawkins, M. E., Knutson, J. R. & Mukerji, I. (2006). Examination of the premelting transition of DNA A-tracts using a fluorescent adenosine analogue. *Biochemistry*, **45**, 5039–5047.
  41. Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. (1987). The structure of an oligo(dA)-oligo(dT) tract and its biological implications. *Nature*, **330**, 221–226.
  42. Coll, M., Frederick, C. A., Wang, A. H.-J. & 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. USA*, **84**, 8385–8389.
  43. Basu, S., Rambo, R. P., Strauss-Soukup, J., Cate, J. H., Ferre-D’Amare, A. R., Strobel, S. A. & Doudna, J. A. (1998). A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nat. Struct. Biol.* **5**, 986–992.
  44. Villeret, V., Huang, S., Fromm, H. J. & Lipscomb, W. N. (1995). Crystallographic evidence for the action of potassium, thallium, and lithium ions on fructose-1,6-bisphosphatase. *Proc. Natl Acad. Sci. USA*, **92**, 8916–8920.
  45. Wulfsberg, G. (1991). *Principles of Descriptive Inorganic Chemistry*. University Science Books, Sausalito, CA.
  46. Morth, J. P., Pedersen, B. P., Toustrup-Jensen, M. S., Sorensen, T. L., Petersen, J., Andersen, J. P. *et al.* (2007). Crystal structure of the sodium-potassium pump. *Nature*, **450**, 1043–1049.
  47. Howerton, S. B., Nagpal, A. & Williams, L. D. (2003). Surprising roles of electrostatic interactions in DNA-ligand complexes. *Biopolymers*, **69**, 87–99.
  48. Howerton, S. B., Sines, C. C., VanDerveer, D. & Williams, L. D. (2001). Locating monovalent cations in the grooves of B-DNA. *Biochemistry*, **40**, 10023–10031.

49. Williams, L. D. & Maher, L. J. (2000). Electrostatic mechanisms of DNA deformation. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 497–521.
50. Moulaei, T., Maehigashi, T., Lountos, G. T., Komeda, S., Watkins, D., Stone, M. P. *et al.* (2005). Structure of B-DNA with cations tethered in the major groove. *Biochemistry*, **44**, 7458–7468.
51. Hud, N. V., Schultze, P., Sklenar, V. & Feigon, J. (1999). Binding sites and dynamics of ammonium ions in a telomere repeat DNA quadruplex. *J. Mol. Biol.* **285**, 233–243.
52. Sandman, K., Krzycki, J. A., Dobrinski, B., Lurz, R. & Reeve, J. N. (1990). Hmf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermobacter feroidus*, is most closely related to histones. *Proc. Natl Acad. Sci. USA*, **87**, 5788–5791.
53. Bailey, K. A. & Reeve, J. N. (1999). DNA repeats and archaeal nucleosome positioning. *Res. Microbiol.* **150**, 701–709.
54. Swinger, K. K. & Rice, P. A. (2004). IHF and Hu: flexible architects of bent DNA. *Curr. Opin. Struct. Biol.* **14**, 28–35.
55. Swinger, K. K. & Rice, P. A. (2007). Structure-based analysis of HU–DNA binding. *J. Mol. Biol.* **365**, 1005–1016.
56. Anderson, J. E., Ptashne, M. & Harrison, S. C. (1987). Structure of the repressor–operator complex of bacteriophage 434. *Nature*, **326**, 846–852.
57. Koudelka, G. B., Harrison, S. C. & Ptashne, M. (1987). Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and Cro. *Nature*, **326**, 886–888.
58. Mauro, S. A. & Koudelka, G. B. (2004). Monovalent cations regulate DNA sequence recognition by 434 repressor. *J. Mol. Biol.* **340**, 445–457.
59. Chen, S., Vojtechovsky, J., Parkinson, G. N., Ebright, R. H. & 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.
60. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature*, **365**, 512–520.
61. DeDecker, B. S., O'Brien, R., Fleming, P. J., Geiger, J. H., Jackson, S. P. & Sigler, P. B. (1996). The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *J. Mol. Biol.* **264**, 1072–1084.
62. Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder, R. G. & Burley, S. K. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl Acad. Sci. USA*, **93**, 4862–4867.
63. Faiger, H., Ivanchenko, M. & 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.
64. Gabrielsen, O. S., Sentenac, A. & Fromageot, P. (1991). Specific DNA binding by c-Myb: evidence for a double helix-turn-helix-related motif. *Science*, **253**, 1140–1143.
65. Schwabe, J. W., Chapman, L., Finch, J. T. & 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.
66. Hegde, R. S., Grossman, S. R., Laimins, L. A. & 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*, **359**, 505–512.
67. Blakaj, D. M., Kattamuri, C., Khrapunov, S., Hegde, R. S. & Brenowitz, M. (2006). Indirect readout of DNA sequence by papillomavirus E2 proteins depends upon net cation uptake. *J. Mol. Biol.* **358**, 224–240.
68. Donner, A. L., Paa, K. & Koudelka, G. B. (1998). Carboxyl-terminal domain dimer interface mutant 434 repressors have altered dimerization and DNA binding specificities. *J. Mol. Biol.* **283**, 931–946.
69. Donner, A. L., Carlson, P. A. & Koudelka, G. B. (1997). Dimerization specificity of P22 and 434 repressors is determined by multiple polypeptide segments. *J. Bacteriol.* **179**, 1253–1261.
70. Woods, K., Lan, T., McLaughlin, L. W. & Williams, L. D. (2003). The role of minor groove functional groups in DNA hydration. *Nucleic Acids Res.* **31**, 1536–1540.
71. Meena, Sun, Z., Mulligan, C. & McLaughlin, L. W. (2006). Removal of a single minor-groove functional group eliminates A-tract curvature. *J. Am. Chem. Soc.* **128**, 11756–11757.
72. Taylor, R., Kennard, O. & Versichel, W. (1984). Geometry of the N–H–O=C hydrogen bond. 2. Three-center (“bifurcated”) and four-center (“trifurcated”) bonds. *J. Am. Chem. Soc.* **106**, 244–248.
73. Dervan, P. (1986). Design of sequence-specific DNA binding molecules. *Science*, **232**, 464–471.
74. Tullius, T. D. & Greenbaum, J. A. (2005). Mapping nucleic acid structure by hydroxyl radical cleavage. *Curr. Opin. Chem. Biol.* **9**, 127–134.
75. Goodsell, D. S., Kaczor-Grzeskowiak, M. & Dickerson, R. E. (1994). The crystal structure of C–C–A–T–T–A–A–T–G–G. Implications for bending of B-DNA at T–A steps. *J. Mol. Biol.* **239**, 79–96.
76. Coulondre, C. & 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.
77. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology, Volume 276: Macromolecular Crystallography* (Carter, C. W., Jr & Sweet, R. M., eds), pp. 307–326, Academic Press, New York; part A.
78. Bailey, S. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–763.
79. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
80. Perrakis, A., Harkiolaki, M., Wilson, K. S. & Lamzin, V. S. (2001). ARP/wARP and molecular replacement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **57**, 1445–1450.
81. Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). Refmac5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2184–2195.
82. Crick, F. H. C. & Magdoff, B. S. (1956). The theory of isomorphous replacement for protein crystals. *Acta Crystallogr.* **9**, 901–908.
83. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.

- 
84. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54**, 905–921.
85. Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G. & Baker, N. A. (2007). PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522–W525.