Published online 27 November 2014 *Nucleic Acids Research, 2014, Vol. 42, No. 22* ***14053–14059***
doi: 10.1093/nar/gku1259

Specific minor groove solvation is a crucial determinant of DNA binding site recognition

Lydia-Ann Harris[1], Loren Dean Williams[2] and Gerald B. Koudelka[1,*]

[1]Department of Biological Sciences, 607 Cooke Hall, University at Buffalo, Buffalo, NY 14260, USA and [2]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA

Received August 29, 2014; Revised November 03, 2014; Accepted November 14, 2014

ABSTRACT

The DNA sequence preferences of nearly all sequence specific DNA binding proteins are influenced by the identities of bases that are not directly contacted by protein. Discrimination between non-contacted base sequences is commonly based on the differential abilities of DNA sequences to allow narrowing of the DNA minor groove. However, the factors that govern the propensity of minor groove narrowing are not completely understood. Here we show that the differential abilities of various DNA sequences to support formation of a highly ordered and stable minor groove solvation network are a key determinant of non-contacted base recognition by a sequence-specific binding protein. In addition, disrupting the solvent network in the non-contacted region of the binding site alters the protein's ability to recognize contacted base sequences at positions 5–6 bases away. This observation suggests that DNA solvent interactions link contacted and non-contacted base recognition by the protein.

INTRODUCTION

Protein recognition of specific DNA sequences is a prerequisite for biological transactions ranging from control of gene expression to DNA restriction and modification and regulation of chromatin structure. Proteins achieve sequence specificity by recognizing both directly contacted bases and subsets of bases in binding sites that are not contacted. Computational (1,2) and empirical studies (e.g. 3–7) show that the relative contributions of non-contacted and contacted base recognition to the overall binding energies vary among protein–DNA complexes. Nonetheless, both components are significant features of nearly all sequence-specific protein–DNA interactions.

Proteins recognize contacted bases via direct molecular interactions between functional groups of amino acid side chains or protein backbone atoms and DNA bases. This so-called 'direct readout' strategy relies on complementary electrostatics, hydrogen bond donors/acceptors, hydrophobic surfaces and the matching of DNA and protein shapes at the interface. Non-contacted base recognition depends on sequence-specific deviations from ideal B-DNA. Proteins recognize non-contacted bases by exploiting DNA sequence-dependent differences in DNA conformation, conformational polymorphism and/or deformability. The structural and physicochemical contributors to non-contacted base recognition by proteins are not completely known.

To help delineate the precise structural mechanisms that govern DNA conformation and conformational flexibility in non-contacted base recognition, we are exploring the DNA binding specificity determinants of bacteriophage P22 repressor (P22R) protein. To accurately regulate the genes needed for establishment and maintenance of lysogeny, P22R must bind and discriminate between six, highly similar, rotationally symmetric DNA binding site sequences in the phage chromosome (Figure 1A). Amino acid side chains in the helix-turn-helix unit of each monomer of the P22R dimer directly contact the outermost bases in one-half of the binding site on the DNA (Figure 1B). The sequences of the contacted bases in each of the six naturally occurring P22R DNA binding sites are nearly identical (Figure 1A). X-ray structural analyses show that no protein atoms closely approach any of the major or minor groove functional groups on the bases at the center of the P22R binding site (8–10), meaning that P22R discriminates between these similar sequence binding sites by indirectly reading the sequence of non-contacted bases at the center of the P22R binding site.

P22R binding induces a transition of the DNA in the non-contacted region of its binding site from B-form to the B′-form (8,9). B′-form DNA is characterized by a much narrower minor groove than that in B-form DNA. The narrow B′ minor groove contains a highly ordered, multi-layered network of solvent molecules. The B′ form is preferred in DNA sequences containing contiguous $^{5'}$ApA$^{3'}$

*To whom correspondence should be addressed. Tel: +1 716 645 4940; Fax: +1 716 645 2975; Email: koudelka@buffalo.edu
Present address: Lydia-Ann Harris, Washington University School of Medicine, Division of Nutritional Science, Campus Box 8031, 660 S. Euclid Avenue, St. Louis, MO 63110, USA.

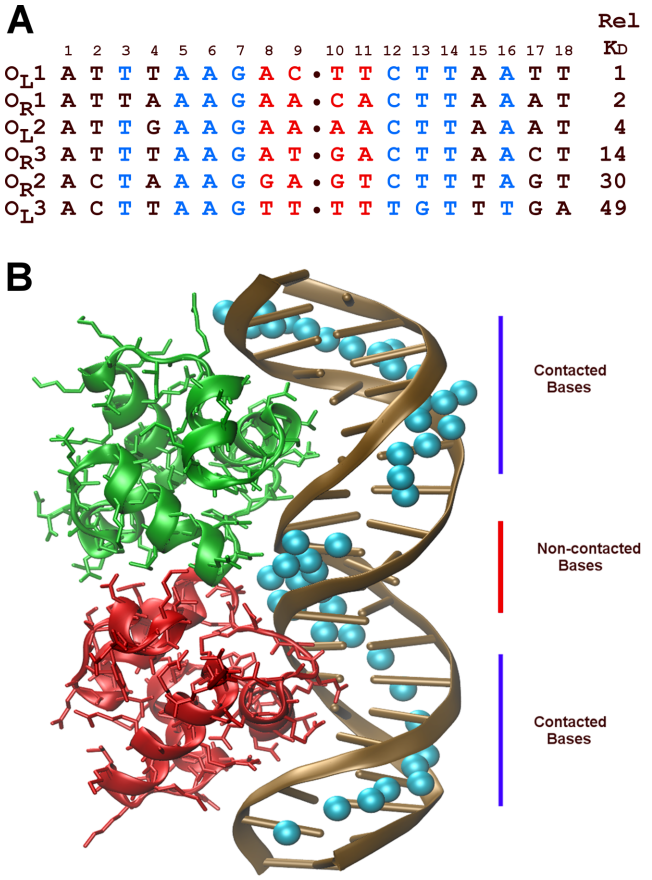

Figure 1. (**A**) P22 operator sequences, the sequences of the naturally occurring P22 operators. Bases highlighted in blue represent sequences in the naturally occurring operators that are directly contacted by the P22R (8,9). Note high degree of sequence conservation among these bases in the various operators. Bases highlighted in red blue denote bases in the non-conserved sequences at the center of the P22 operators that are not contacted by P22R (8,9). The center of symmetry of all the sequences lies between bases 9 and 10 and is highlighted by the 'closed circle'. The 'numbers' denote the relative dissociation constants (K_D) of these operators for P22 repressor as initially reported in (31). The values are normalized to the K_D of O_L1, the tightest binding natural operator, which is 5×10^{-9} M. (**B**) P22R NTD-DNA^{9T} complexes. The DNA binding domain of P22R (ribbon) binds as a dimer to an 18 base pair consensus sequence. Protein residues in the helix-turn-helix unit in each monomer make direct contacts with the major groove in outer half of the DNA binding site. P22R does not contact the central four base pairs (indicated). The DNA bases are rendered as sticks, and the DNA backbone is rendered as a tube. DNA-associated solvent molecules (i.e. oxygens of water or ions) are rendered as spheres.

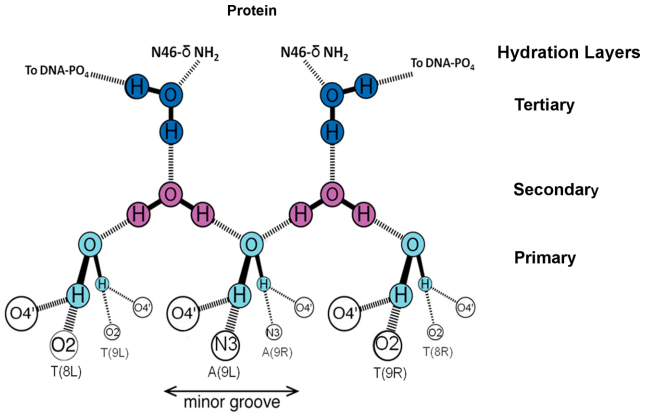

Figure 2. Model of the minor groove solvation network. Primary water molecules, which directly contact DNA functional groups, are colored cyan, secondary waters pink and tertiary waters blue. Predicted hydrogen bonds formed between water molecules, DNA phosphates, minor groove functional groups and protein amino acid side chains are depicted by broken lines.

and/or ApT steps (e.g. A-tracts (11–14)) and can be induced in A/T-rich sequences (15,16). However prior to our work with P22R (9), G/C-containing DNAs were not found in the B′-form. Crystallographic evidence indicates P22R binding induces a slight bend in the binding site, but the degree of induced bend does not vary with non-contacted base sequence. Spectroscopic evidence indicates that the B′-form is imposed on the non-contacted bases in DNA by binding of P22R (17). The juxtaposition of four negatively charged amino acids near DNA phosphates is an important mechanistic driver favoring the B′ state (9,17). The varying propensity of different sequences to assume the B′ form (17) ($^{5'}$ApA$^{3'} \sim {}^{5'}$ApT$^{3'} > {}^{5'}$TpA$^{3'} > {}^{5'}$CpG$^{3'} \sim {}^{5'}$GpC$^{3'}$ (18)) helps P22R to discriminate between binding sites bearing different non-contacted bases. The features that govern the varying propensities of different DNA sequences to transition to the B′ form have not been delineated.

P22R binding induces a B → B′ transition in the non-contacted region of the binding site, irrespective of the non-contacted base sequence and base identity (9,10). However, the minor groove solvation pattern in the induced B′ region varies with its sequence. The floor of the minor groove of A/T sequences contains only H-bond acceptors (Figure 2). In P22R complexes with DNAs containing ApT or TpA in the non-contacted region, the waters in the primary solvent layer simultaneously contribute two hydrogens to four H-bond acceptors in DNA. The waters in this layer also accept H-bonds from water molecules in the second solvent layer. The waters in the secondary layer accept hydrogen bonds from solvent in the tertiary layer. As a consequence of this multi-layered structure, the solvent molecules in the primary layer are highly constrained to discrete minor groove sites. In G/C-containing P22R–DNA complexes, the presence of the H-bond accepting 2-NH_2 group in the minor groove of the non-contacted bases allows the primary water molecules to interact with closely spaced sites lined by a mixture of hydrogen bond donors and acceptors, releasing the translational and rotational restraints of the solvent. Expected sequence-dependent differences in the mobility of these primary water molecules are observable in chemical footprinting experiments (9). These observations suggest that the ability of a particular DNA sequence to support the formation of a highly ordered and stable minor groove solvation network is a key determinant of non-contacted base recognition.

MATERIALS AND METHODS

Binding sites, purification and end labeling

DNA oligonucleotides were obtained from Integrated DNA Technologies. Each DNA encodes only a single P22R

binding site. Oligonucleotides were purified from denaturing polyacrylamide gels containing 7-M urea, 8% polyacrylamide and TBE (89-mM Tris pH 7.5 and 89-mM boric acid, 1-mM ethylenediaminetetraacetic acid (EDTA)). After desalting on PD-10 columns (GE Lifesciences, Piscataway, NJ, USA) in TEN (10-mM Tris pH 7.5, 50-mM NaCl, 1-mM EDTA), equimolar quantities of complementary single strands were annealed by heating to 80°C for 2 min and slow-cooling to room temperature. Double-stranded DNAs were separated from single strands by gel electrophoresis on 8% polyacrylamide gels and purified as described above using an 8% native acrylamide gel.

For filter binding experiments, 1 μg of DNA was 5′ end labeled with adenosine 5' triphosphate [γ-32P] (6000 Ci/mmol) (Perkin-Elmer, Boston, MA, USA) in the presence of T4 Polynucleotide Kinase (Epicentre, Inc., Madison, WI, USA). The resulting products were ethanol precipitated from a 0.5-M ammonium acetate solution and resuspended in TE (Tris pH 7.5, 1-mM EDTA). The 9T-(3C) DNA fragment was prepared and purified as described (19).

Preparation of P22R mutants

Plasmids directing overproduction of E44A, N46C, N46S and N46A were constructed by site directed mutagenesis using the plasmid pTPΔ125 (20,21) as a template. Primers used for mutagenesis were obtained from Integrated DNA Technologies. Wild-type (WT) and mutant P22 repressors were isolated from the *Escherichia coli* strain X90 (22) bearing the appropriate plasmids. All mutant proteins were purified as described by DeAnda *et al.* (20). The activity of all proteins was determined as described (23).

Filter binding assay

The dissociation constants (K_D) of wild-type (WT) and mutant P22 repressor–DNA complexes were measured by filter binding. These experiments were performed as described previously (24). We evaluated the binding of WT and P22 repressor variants to various synthetic sites in a reaction buffer composed of 100-mM Tris, 50-mM NaCl, 0.1-mg/ml bovine serum albumin, 10-mM IPTG and 1-mM dithiothreitol. The concentration of DNA used in each experiment was <0.27 nM. After filtering each sample through stacked nitrocellulose and DEAE membranes (25), and exposure to a Phosphorimaging screen, counts retained on each filter were quantified using Image Quant 5.0 software. Dissociation constants were determined by non-linear, least squares fitting of the filter binding data to a hyperbolic equation using Prism 3.0 software (GraphPad Software Inc.). DNAs contain only a single binding site, which eliminates the need to account effect of cooperative binding interactions between two DNA-bound P22 repressors on P22R DNA binding affinity. Each K_D was determined from 10 to 15 replicate measurements. Student's two-tailed, two-sample equal variance *t*-tests were used to determine the significance of comparisons between measured data.

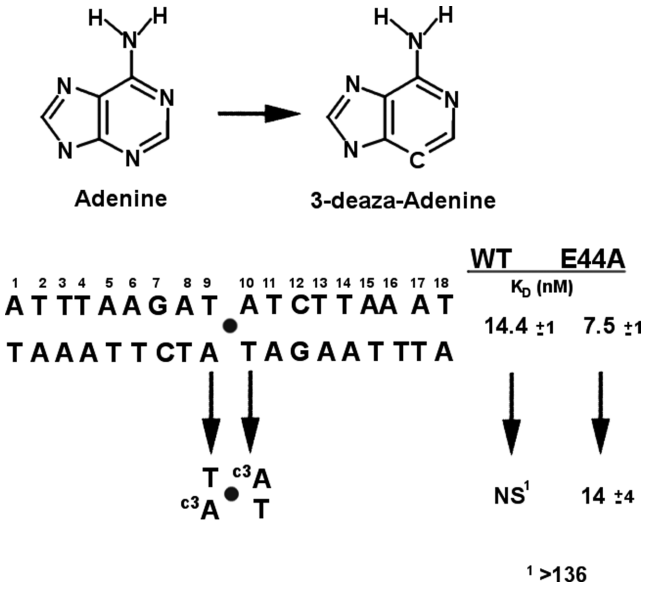

Figure 3. Effect of removing minor groove functional groups from non-contacted bases on the DNA affinity of wild-type and E44A mutant P22R. The adenines at the symmetrically related positions 9 and 10 were substituted with 3-deaza-adenine (9^{c3}A). Shown is the sequence of both strands of the synthetic P22R DNA binding site, DNA^{9T}. The dissociation constants of wild-type and E44A mutant P22R complexes with DNA^{9T} and DNA^{9c3A} are given ± standard deviation.

RESULTS

Role of minor groove solvent network in non-contacted base recognition

We examined the importance of the minor groove solvent network to P22R DNA binding by comparing the affinities of WT and mutant P22R for DNAs bearing either adenine at the non-contacted positions 9 and 10 (DNA^{9T}) or the non-canonical 3-deaza-adenine base at positions at these positions (DNA^{9Tc3A}) (Figure 3). The N3 of adenines at positions 9 and 10 in the binding site accepts H-bonds from the primary solvent molecule in the first layer of the solvation spine (Figure 1). Replacing the N3 atoms on adenines 9 and 10 with atoms that are incapable of forming H-bonds should disorder the solvent network in the minor groove. We find that WT P22R binds strongly to DNA^{9T} (Figure 3). In contrast, P22R does not bind specifically to DNA^{9Tc3A}; its affinity for DNA^{9Tc3A} is indistinguishable from its affinity for non-specific DNA (Figure 3). This finding suggests that assembly of a specific geometric arrangement of the solvent molecules in the minor groove of the non-contacted bases is a prerequisite for P22R–DNA complex formation.

In addition to affecting water and ion binding to DNA, replacing the N3 atom on adenine with carbon can slightly destabilize DNA double helices (26). To confirm that the 3-deaza-adenine substitution impacts P22R DNA binding by affecting DNA-solvent interactions and not by destabilizing dsDNA, we examined the ability of E44A mutant P22R to bind DNA^{9T} and $DNA^{9T\text{-}c3A}$. The DNA affinities of P22R mutants lacking a negatively charged residue at position 44 are less sensitive changes in non-contact base sequence than are those of WT P22R (17). In contrast to WT P22R, the

P22R E44A mutant binds tightly to DNA^{9T} and DNA^{9Tc3A} sites, with similar affinities (Figure 3). This finding indicates that any putative DNA destabilization caused by 3-deaza-adenine substitution is at most a minor contributor to differential binding of WT P22R. Together, the results in Figure 3 indicate that the presence of the N3 H-bond accepting group on the minor groove floor in the non-contacted region is essential for sequence-specific binding by P22R. More importantly, these observations support the idea that non-contacted base recognition by P22R is mediated by the ability of these bases to support formation of an organized solvent network in the minor groove.

In P22R–DNA complexes, asparagine 46 (N46) is located at the mouth of minor groove in the central (non-contacted) region of the binding site. The δ-NH_2 of N46 forms H-bonds with water molecules in the tertiary layer of the minor groove solvent network (Figure 2). In this position, N46 may help organize and/or stabilize the solvent network in the minor groove of non-contacted base sequences. To test this model, we determined the affinities of mutant P22Rs bearing N46A, N46C or N46S substitutions for three binding sites, DNA^{9A}, DNA^{9T} and DNA^{9C}. These binding sites vary in their propensity to form B′ structure, with the A-tract containing DNA^{9A} having a 'pre-formed' B′ structure in its non-contacted base region and DNA^{9T} and DNA^{9C} displaying increasing resistance to forming this structure (9,17). To facilitate comparisons between the affinities of WT P22R and N46x mutant proteins for the various binding sites, we performed the binding experiments at lower salt concentration (50-mM NaCl). Under this condition, WT P22R binds the DNA^{9A} site >3-fold better than it does to DNA^{9T} and ∼13-fold better than to DNA^{9C} (Table 1). The N46C and N46S P22R variants, which bear H-bond forming amino acids at position 46, are able to bind all three sites, DNA^{9A} DNA^{9T} and DNA^{9C}, albeit with lowered discrimination abilities and affinities than does WT P22R. In contrast to these proteins, the N46A mutant protein is completely incapable of specifically binding to either DNA^{9C} or DNA^{9T} (Table 1). Nonetheless, the N46A mutant specifically binds the A-tract containing DNA^{9A}. Moreover, its affinity for this site is similar to that of WT P22R. These results demonstrate that a hydrogen bonding amino acid side chain at position 46 is required for P22R binding to DNAs bearing non-contacted base sequences that do not naturally assume the B′ form. This observation suggests that the residue at position 46 plays a crucial role in inducing and/or maintaining the organized solvent network in the minor groove of the non-contacted bases. More importantly, these results provide further evidence that minor groove solvation is the key determinant of non-contacted base recognition by P22R.

Minor groove solvation: coupling of non-contacted and contacted base recognition

Previous work indicated an intriguing linkage between P22R recognition of non-contacted base sequence and its direct readout of the contacted bases. That is, sequence changes in the non-contacted bases alter P22 repressor's contacted base preferences (19). To explore how P22R's contacted and non-contacted base recognition mechanisms

Table 1. Role of a hydrogen bonding residue at position 46 in modulating non-contacted base recognition by P22R

Protein	9T	9C	9A
		K_D (nM)	
Wild-type	14.4	57.6	4.3
N46A	NS[a]	NS[a]	2.8
N46S	50.4	98	47.5
N46C	57.6	78	25.9

[a] >345 nM.

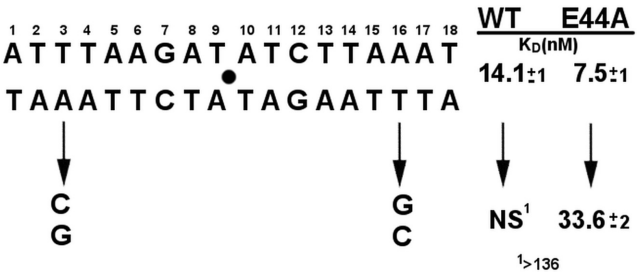

Figure 4. Effect of a changing contacted position 3 on the DNA affinity of wild-type and E44A mutant P22R. Shown is the sequence of both strands of the synthetic P22R DNA binding site, DNA^{9T}. The bases at the symmetrically related positions 3 and 16 in the P22R DNA binding site were changed from T•A to C•G. Shown are the dissociation constants of wild-type and E44A mutant P22R complexes with DNA^{9T} and DNA$^{9T(3C)}$ are given ± standard deviation.

are linked, we first determined the affinities of WT P22R and E44A repressors for DNA$^{9T(3C)}$, a binding site that contains T•A→C•G substitutions at the symmetrically related, contacted position 3 at each end of the binding site. WT P22R binds DNA$^{9T(3C)}$ only non-specifically with a K_D ≥136 nM, an affinity that is ≥16-fold lower than its affinity for DNA^{9T} (Figure 4). In contrast, P22R-E44A binds DNA$^{9T(3C)}$ with an affinity <5-fold lower than its affinity for DNA^{9T}. This result shows that E44A P22R, which has a reduced ability to distinguish between changes in non-contacted base sequence, also has reduced ability to distinguish between base changes at a contacted position. These observations confirm that P22R recognition of contacted and non-contacted bases is linked.

The sequences of the symmetrically related contacted bases at positions 2–5 and 14–17 comprise A-tracts. As expected (11,12), the minor grooves of these bases contain a highly organized, multilayer solvent network (8).To test whether this solvent network plays a role in linking contacted and non-contacted base recognition, we examined the affinity of WT P22R and E44A proteins for binding sites bearing 3-deaza-adenine symmetrically substituted at positions 2 and 3 (DNA$^{9T(2c3A\text{-}3c3A)}$ at each end of the binding site (Figure 5)). This substitution removes H-bond acceptors from the minor groove surfaces of these bases without changing any of the major groove atoms on these bases that are directly contacted by P22R. Intriguingly, WT P22R does not bind specifically to DNA$^{9T(2c3A\text{-}3c3A)}$ (Figure 5). This finding indicates that the presence of the N3 H-bond accepting groups on the minor groove surface of the contacted bases is essential for DNA complex formation by P22R. In contrast to WT P22R, the E44A mutant repres-

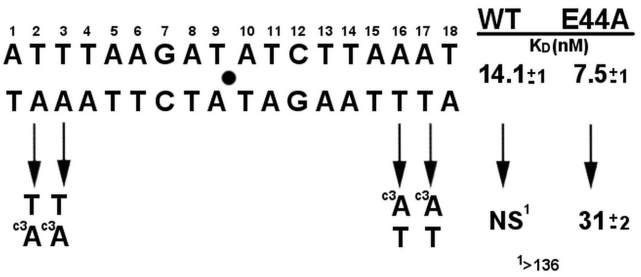

Figure 5. Effect of removing minor groove functional groups from contacted bases on the DNA affinity of wild-type and E44A mutant P22R. Shown is the sequence of both strands of the synthetic P22R DNA binding site, DNA^{9T}. The bases at the symmetrically related positions 2 and 3 and 16 and 17 were substituted with 3-deaza-adenine (^{c3}A). Shown are the dissociation constants of wild-type and E44A mutant P22R complexes with DNA^{9T} and $DNA^{9T(2c3A\text{-}3c3A)}$ are given ± standard deviation.

sor binds $DNA^{9T(2c3A\text{-}3c3A)}$ with an affinity only 4-fold lower than its affinity for DNA^{9T}. This result echoes the finding that non-contacted base recognition affects P22R's recognition of contacted bases (Figure 4). More importantly, these results suggest the connection between contacted and non-contacted base recognition relies on an organized minor groove solvent network in the contacted region of the P22R binding site.

DISCUSSION

We previously suggested that a key determinant for non-contacted base recognition, and thus binding site discrimination by P22R, is the propensity of various base sequences to assume the B′-DNA state (9–10,17). This B→B′ transition is characterized by substantial narrowing of the minor groove from 12 Å→2.2 Å and formation of a highly organized stacked monolayer of rotationally and translationally restrained water molecules (8–10). Here we show that a P22R binding site containing a 'pre-formed', B′-like solvent network in the minor groove strongly binds P22R. In contrast, sequences and modified bases that cannot support formation of an organized minor groove solvent network either do not bind P22R or bind it poorly. These results argue that discrimination between non-contacted base sequences by P22R is governed substantially by the differential abilities of various sequences to support formation of a B′-like solvent network in the minor groove.

It is unclear whether formation of the minor groove solvent network facilitates the B→B′ transition of the non-contacted bases in P22R–DNA complexes or is needed to stabilize the B′ form. However, (i) the non-contacted regions of most of the natural and synthetic P22R binding sites studied (17,27) are not likely to be in the B′ form in the absence of P22R and; (ii) our evidence strongly indicates P22R binding induces the B→B′ transition (17). Additionally, although all solvent peaks in the P22R–DNA complexes are assigned as water molecules, it is likely that some positions are at least partially occupied by cations (28–30). Consistent with this suggestion, biochemical studies indicate involvement of cations in indirect readout of non-contacted bases of P22R (19,21,27,31). The presence of cations in the solvent network may stabilize the extreme narrowing of the minor groove in the P22R–DNA complexes. Together, these observations suggest that formation of the minor groove solvent network is a consequence of the P22R-induced B→B′ transition.

Two non-contacted base recognition modes have been described: indirect readout (32) and shape recognition (33). However these seemingly distinct recognition mechanisms are apparently influenced by the same underlying DNA feature—a narrow minor groove. Indirect readout depends on the propensity of a given sequence to assume a conformation that facilitates protein binding, i.e. one that contains a very narrow groove in the non-contacted base region of the respective protein–DNA complexes (4,7,9,17,34–41). This narrow groove is required to bring the two halves of the DNA binding site into proper alignment with the bound protein. In shape recognition, DNA sequence-dependent variation of minor groove width locally alters the negative electrostatic potential of DNA via electrostatic focusing (42–46). The corresponding variations in electrostatic potential are recognized by positively charged protein residue(s) to read the 'shape' (width) of the DNA minor groove. Hence in both cases, protein binding to DNA is ultimately governed by the propensity of a given non-contacted base DNA sequence to support formation of a narrow minor groove (9,17,40,42,47–48).

Since indirect readout and shape recognition would both appear to depend on minor groove width, the question becomes; do similar DNA structural features govern minor groove width in both mechanisms? The structure and dynamic polymorphism of DNA is governed by sequence-dependent intrinsic (e.g. base stacking interactions (49)), steric repulsion among exocyclic groups on the bases (50) and extrinsic forces (i.e. electrostatic interactions of DNA with its environment (30,46)). We and others showed that the 2-NH_2 group in the minor groove of the non-contacted bases is the primary mediator of indirect readout (7,17,40). When present in non-contacted bases, this functional group, which is found in G/C-containing DNA, substantially destabilizes protein–DNA complexes. Similarly, proteins that rely on shape recognition for binding site discrimination also bind weakest to sites containing G/C-rich non-contacted regions and most tightly to DNAs with A/T-rich non-contacted DNA sequences. Hence sequence-dependent variation in minor groove width depends on the composition of functional groups in the minor groove of the non-contacted bases, specifically the absence or presence of a 2-NH_2 group.

How does the 2-NH_2 group affect minor groove width? Our findings challenge long-held ideas of how differences in DNA sequence lead to variation in DNA conformation/conformational flexibility. It has been long suggested that sequence-dependent differences in DNA conformation arise from (i) differences in the hydrogen bonding pattern of each base pair, (ii) from differing stacking interactions for each dinucleotide step; and (iii) the need to minimize cross-strand steric clashes between exocyclic groups, including the 2-NH_2 group (2,51–54), e.g. via changes in base pair roll (2,7,55–57). The need to optimize these types of 'intrinsic' intramolecular interactions is therefore predicted to lead to the sequence-dependent heterogeneity in DNA structure and minor groove width.

The 2-NH_2 on guanine has also been thought to provide a steric and electrostatic impediment to groove collapse (58). In contrast to these ideas, our data indicate that the tendency of the minor groove to narrow depends on the ability of a stable, organized solvent network to assemble in this groove. Our findings indicate that minor groove width is governed by the structure and stability of this solvent network, as modulated by the precise composition of minor groove functional groups. We note that for the ETS domain protein PU.1, a well-ordered hydration structure in the major groove is necessary for the formation of its high-affinity complex with DNA (59).

Contacted base recognition is markedly influenced by non-contacted base sequence (Figures 4 and 5). The first contacted base (position 2) in the P22R binding site is 7 base pairs away from the non-contacted region. These observations indicate recognition of contacted and non-contacted bases by P22R is not physically, but allosterically, linked. Since the conformations of unbound and bound P22R are identical (8,60) and the conformation of the DNA-bound P22R does not vary with non-contacted base sequence, we suggest that the allosteric connection between these two recognition modes is mediated through DNA. Among other roles, 'through-DNA' allostery has been implicated in regulating cooperative interactions between DNA binding proteins and DNA binding of intercalating drugs (61–64). Support for a through-DNA allosteric connection that links contacted and non-contacted base recognition comes from the finding that the DNAse I and •OH cleavage intensity at contacted positions are markedly influenced by the base sequences at the non-contacted positions (19,27) in both unbound and P22R-bound DNA.

A continuous, connected solvent spine runs from the contacted bases at positions 2–7 and 12–17 to the edge of the non-contacted regions (8) (Figure 1A). As such, this network is directly juxtaposed with the solvent network in the minor groove of the non-contacted bases at positions 8–11. Since disruption of either solvent network affects P22R recognition of both contacted and non-contacted bases, we conjecture that the allosteric connection between contacted and non-contacted sequence recognition is mediated by interaction between the solvent networks in the minor grooves of the non-contacted and contacted bases. Regardless of whether these two solvent networks directly interact, our results indicate that formation of a highly ordered and stable minor groove solvation network is a key determinant of contacted and non-contacted base recognition by P22R.

FUNDING

Funding for open access charge: University at Buffalo [to G.K.].
Conflict of interest statement. None declared.

REFERENCES

1. Michael Gromiha,M., Siebers,J.G., Selvaraj,S., Kono,H. and Sarai,A. (2004) Intermolecular and intramolecular readout mechanisms in protein–DNA recognition. *J. Mol. Biol.*, **337**, 285–294.
2. Rohs,R., Jin,X., West,S.M., Joshi,R., Honig,B. and Mann,R.S. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.*, **79**, 233–269.
3. Ferreiro,D.U., Dellarole,M., Nadra,A.D. and de Prat-Gay,G. (2005) Free energy contributions to direct readout of a DNA sequence. *J. Biol. Chem.*, **280**, 32480–32484.
4. 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.
5. Chen,S., Gunasekera,A., Zhang,X., Kunkel,T.A., Ebright,R.H. and Berman,H.M. (2001) Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: alteration of DNA binding specificity through alteration of DNA kinking. *J. Mol. Biol.*, **314**, 75–82.
6. Gunasekera,A., Ebright,Y.W. and Ebright,R.H. (1992) DNA sequence determinants for binding of the Escherichia coli catabolite gene activator protein. *J. Biol. Chem.*, **267**, 14713–14720.
7. Hancock,S.P., Ghane,T., Cascio,D., Rohs,R., Di Felice,R. and Johnson,R.C. (2013) Control of DNA minor groove width and Fis protein binding by the purine 2-amino group. *Nucleic Acids Res.*, **41**, 6750–6760.
8. Watkins,D., Hsiao,C., Woods,K.K., Koudelka,G.B. and Williams,L.D. (2008) P22 c2 repressor:operator complex: mechanisms of direct and indirect readout. *Biochemistry*, **47**, 2325–2338.
9. Watkins,D., Mohan,S., Koudelka,G.B. and Williams,L.D. (2010) Sequence recognition of DNA by protein-induced conformational transitions. *J. Mol. Biol.*, **396**, 1145–1164.
10. Watkins,D., Harris,L., Koudelka,G.B. and Williams,L.D. (2011) Protein: Feeling the Groove of DNA. In: Sheardy,R. D. and Winkle,S. A. (eds). *Frontiers in Nucleic Acids*. American Chemical Society, Washington, DC, pp. 147–166.
11. 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.
12. Shui,X., Sines,C.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.
13. Sunnerhagen,M., Denisov,V.P., Venu,K., Bonvin,A.M., Carey,J., Halle,B. and Otting,G. (1998) Water molecules in DNA recognition I: hydration lifetimes of trp operator DNA in solution measured by NMR spectroscopy. *J. Mol. Biol.*, **282**, 847–858.
14. Liepinsh,E., Leupin,W. and Otting,G. (1994) Hydration of DNA in aqueous solution: NMR evidence for a kinetic destabilization of the minor groove hydration of d-(TTAA)2 versus d-(AATT)2 segments. *Nucleic Acids Res.*, **22**, 2249–2254.
15. 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.
16. Tevis,D.S., Kumar,A., Stephens,C.E., Boykin,D.W. and Wilson,W.D. (2009) Large, sequence-dependent effects on DNA conformation by minor groove binding compounds. *Nucleic Acids Res.*, **37**, 5550–5558.
17. Harris,L.A., Watkins,D., Williams,L.D. and Koudelka,G.B. (2013) Indirect readout of DNA sequence by p22 repressor: roles of DNA and protein functional groups in modulating DNA conformation. *J. Mol. Biol.*, **425**, 133–143.
18. Hud,N.V. and Plavec,J. (2003) A unified model for the origin of DNA sequence-directed curvature. *Biopolymers*, **69**, 144–158.
19. Hilchey,S.P. and Koudelka,G.B. (1997) DNA-based loss of specificity mutations. *J. Biol. Chem.*, **272**, 1646–1653.
20. DeAnda,J., Poteete,A.R. and Sauer,R.T. (1983) P22 c2 repressor-domain structure and function. *J. Biol. Chem.*, **258**, 10536–10542.
21. 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.
22. Coulandre,C. and Miller,J.H. (1977) Genetic studies of the lac repressor. III. Additional correlation of mutational sites with specific amino acid residues. *J. Mol. Biol.*, **117**, 525–567.
23. Riggs,A.D., Suzuki,H. and Bourgeoss,S. (1970) Lac repressor-operator interaction. I. Equilibrium studies. *J. Mol. Biol.*, **48**, 67–83.

24. Mauro,S.A. and Koudelka,G.B. (2004) Monovalent cations regulate DNA sequence recognition by 434 repressor. *J. Mol. Biol.*, **340**, 445–457.
25. Wong,I. and Lohman,T.M. (1993) A double-filter method fornitrocellulose-filter binding—application to protein nucleic-acid interactions. *Proc. Natl Acad. Sci. U.S.A.* **90**, 5428–5432.
26. Ganguly,M., Wang,R.-W., Marky,L.A. and Gold,B. (2010) Thermodynamic characterization of DNA with 3-deazaadenine and 3-methyl-3-deazaadenine substitutions: the effect of placing a hydrophobic group in the minor groove of DNA. *J. Phys. Chem. B*, **114**, 7656–7661.
27. 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**, 9134–9135.
28. Woods,K., McFail-Isom,L., Sines,C.C., Howerton,S.B., Stephens,R.K. and Williams,L.D. (2001) Monovalent cations sequester within the A-tract minor groove of [d(CGCGAATTCGCG)]2. *J. Am. Chem. Soc.*, **122**, 1546–1547.
29. 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.
30. McFail-Isom,L., Sines,C.C. and Williams,L.D. (1999) DNA structure: cations in charge? *Curr. Opin. Struct. Biol.*, **9**, 298–304.
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. Koudelka,G.B., Mauro,S.A. and Ciubotaru,M. (2006) Indirect readout of DNA sequence by proteins: the roles of DNA sequence-dependent intrinsic and extrinsic forces. *Prog. Nucleic Acid Res. Mol. Biol.*, **81**, 143–177.
33. Rohs,R., West,S.M., Liu,P. and Honig,B. (2009) Nuance in the double-helix and its role in protein-DNA recognition. *Curr. Opin. Struct. Biol.*, **19**, 171–177.
34. Stolfa,G. and Koudelka,G.B. (2013) Entry and killing of tetrahymena thermophila by bacterially produced shiga toxin. *mBio*, **4**, e00416-12.
35. Mauro,S.A., Pawlowski,D. and Koudelka,G.B. (2004) The role of the minor groove substituents in indirect readout of DNA sequence by 434 repressor. *J. Biol. Chem.*, **278**, 12955–12960.
36. Koudelka,G.B., Eckstein,F. and Lilley,D.M.J. (1993), *Nucleic Acids and Molecular Biology*. Springer-Verlag, Berlin-Heidelberg, **Vol. 7**, pp. 16–27.
37. Koudelka,G.B. and Carlson,P. (1992) DNA twisting and the effects of non-contacted bases on affinity of 434 operator for 434 repressor. *Nature*, **355**, 89–91.
38. Koudelka,G.B., Harbury,P., Harrison,S.C. and Ptashne,M. (1988) DNA twisting and the affinity of bacteriophage 434 operator for bacteriophage 434 repressor. *Proc. Natl Acad. Sci. U.S.A.*, **85**, 4633–4637.
39. 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*, **326**, 886–888.
40. Hines,C.S., Meghoo,C., Shetty,S., Biburger,M., Brenowitz,M. and Hegde,R.S. (1998) DNA structure and flexibility in the sequence-specific binding of papillomavirus E2 proteins. *J. Mol. Biol.*, **276**, 809–818.
41. Rohs,R., Sklenar,H. and Shakked,Z. (2005) Structural and energetic origins of sequence-specific DNA bending: Monte Carlo simulations of papillomavirus E2-DNA binding sites. *Structure*, **13**, 1499–1509.
42. Joshi,R., Passner,J.M., Rohs,R., Jain,R., Sosinsky,A., Crickmore,M.A., Jacob,V., Aggarwal,A.K., Honig,B. and Mann,R.S. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell*, **131**, 530–543.
43. Slattery,M., Riley,T., Liu,P., Abe,N., Gomez-Alcala,P., Dror,I., Zhou,T., Rohs,R., Honig,B., Bussemaker,H.J. *et al.* (2011) Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell*, **147**, 1270–1282.
44. Dror,I., Zhou,T., Mandel-Gutfreund,Y. and Rohs,R. (2014) Covariation between homeodomain transcription factors and the shape of their DNA binding sites. *Nucleic Acids Res.*, **42**, 430–441.
45. Gordan,R., Shen,N., Dror,I., Zhou,T., Horton,J., Rohs,R. and Bulyk,M.L. (2013) Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep.*, **3**, 1093–1104.
46. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
47. Bishop,E.P., Rohs,R., Parker,S.C., West,S.M., Liu,P., Mann,R.S., Honig,B. and Tullius,T.D. (2011) A map of minor groove shape and electrostatic potential from hydroxyl radical cleavage patterns of DNA. *ACS Chem. Biol.*, **6**, 1314–1320.
48. Locasale,J.W., Napoli,A.A., Chen,S., Berman,H.M. and Lawson,C.L. (2009) Signatures of protein-DNA recognition in free DNA binding sites. *J. Mol. Biol.*, **386**, 1054–1065.
49. Grzeskowiak,K., Yanagi,K., Prive,G.G. and Dickerson,R.E. (1991) The structure of B-helical C-G-A-T-C-G-A-T-C-G and comparison with C-C- A-A-C-G-T-T-G-G. The effect of base pair reversals. *J. Biol. Chem.*, **266**, 8861–8883.
50. Chiu,T.K., Kaczor-Grzeskowiak,M. and Dickerson,R.E. (1999) Absence of minor groove monovalent cations in the crosslinked dodecamer C-G-C-G-A-A-T-T-C-G-C-G. *J. Mol. Biol.*, **292**, 589–608.
51. Klug,A., Jack,A., Viswamitra,M.A., Kennard,O., Shakked,Z. and Steitz,T.A. (1979) A hypothesis on a specific sequence-dependent conformation of DNA and its relation to the binding of the lac-repressor protein. *J. Mol. Biol.*, **131**, 669–680.
52. Gorin,A.A., Zhurkin,V.B. and Olson,W.K. (1995) B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.*, **247**, 34–48.
53. Wartell,R.M., Larson,J.E. and Wells,R.D. (1974) Netropsin: a specific probe for A/T regions of duplex deoxyribonucleic acid. *J. Biol. Chem.*, **249**, 6719–6731.
54. Yoon,C., Prive,G.G., Goodsell,D.S. and Dickerson,R.E. (1988) Structure of an alternating-B DNA helix and its relationship to A-tract DNA. *Proc. Natl Acad. Sci. U.S.A.*, **85**, 6332–6336.
55. Calladine,C.R. (1982) Mechanics of sequence-dependent stacking of bases in B-DNA. *J. Mol. Biol.*, **161**, 343–352.
56. el Hassan,M.A. and Calladine,C.R. (1996) Structural mechanics of bent DNA. *Endeavour*, **20**, 61–67.
57. el Hassan,M.A. and Calladine,C.R. (1996) Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J. Mol. Biol.*, **259**, 95–103.
58. Dror,I., Zhou,T., Mandel-Gutfreund,Y. and Rohs,R. (2014) Covariation between homeodomain transcription factors and the shapeof their DNA binding sites. *Nucleic Acids Res*, **2**, 430–441.
59. Wang,S., Linde,M.H., Munde,M., Carvalho,V.D., Wilson,W.D. and Poon,G.M.K. (2014) Mechanistic heterogeneity in site recognition by the structurally homologous DNA-binding domains of the ETS family transcription factors Ets-1 and PU.1. *J. Biol. Chem.*, **289**, 21605–21616.
60. Sevilla-Sierra,P., Otting,G. and Wüthrich,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.
61. Rudnick,J. and Bruinsma,R. (1999) DNA-protein cooperative binding through variable-range elastic coupling. *Biophys. J.*, **76**, 1725–1733.
62. Kim,S., Broströmer,E., Xing,D., Jin,J., Chong,S., Ge,H., Wang,S., Gu,C., Yang,L., Gao,Y.Q. *et al.* (2013) Probing allostery through DNA. *Science*, **339**, 816–819.
63. Hogan,M., Dattagupta,N. and Crothers,D.M. (1979) Transmission of allosteric effects in DNA. *Nature*, **278**, 521–524.
64. Moretti,R., Donato,L.J., Brezinski,M.L., Stafford,R.L., Hoff,H., Thorson,J.S., Dervan,P.B. and Ansari,A.Z. (2008) Targeted chemical wedges reveal the role of allosteric DNA modulation in protein-DNA assembly. *ACS Chem. Biol.*, **3**, 220–229.