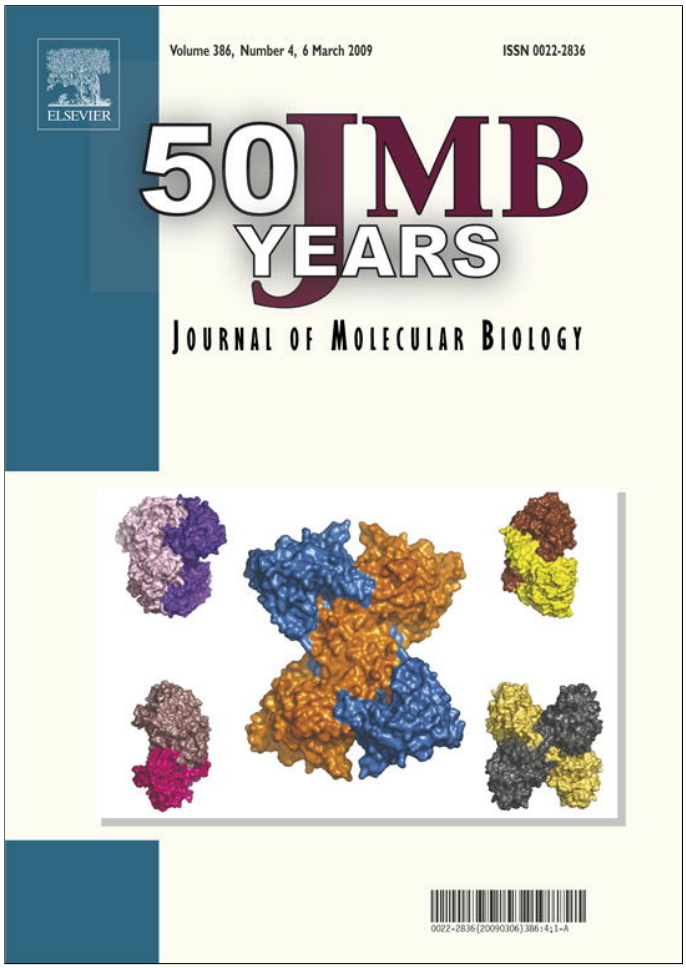

doi:10.1016/j.jmb.2009.01.007 J. Mol. Biol. (2009) 386, 1054–1065

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Signatures of Protein-DNA Recognition in Free DNA Binding Sites

Jason W. Locasale[1,2]†, Andrew A. Napoli[1]†, Shengfeng Chen[1]†, Helen M. Berman[1] and Catherine L. Lawson[1]*

[1]*Department of Chemistry and Chemical Biology, Rutgers, the State University of New Jersey, 610 Taylor Road, Piscataway, NJ 08854, USA*

[2]*Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA*

Received 25 October 2008; received in revised form 30 December 2008; accepted 9 January 2009
Available online
15 January 2009

Edited by P. J. Hagerman

One obstacle to achieving complete understanding of the principles underlying sequence-dependent recognition of DNA is the paucity of structural data for DNA recognition sequences in their free (unbound) state. Here, we carried out crystallization screening of 50 DNA duplexes containing cognate protein binding sites and obtained new crystal structures of free DNA binding sites for three distinct modes of DNA recognition: antiparallel β strands (MetR), helix-turn-helix motif + hinge helices (PurR), and zinc fingers (Zif268). Structural changes between free and protein-bound DNA are manifested differently in each case. The new DNA structures reveal that distinctive sequence-dependent DNA geometry dominates recognition by MetR, protein-induced bending of DNA dictates recognition by PurR, and deformability of DNA along the *A-B* continuum is important in recognition by Zif268. Together, our findings show that crystal structures of free DNA binding sites provide new information about the nature of protein–DNA interactions and thus lend insights towards a structural code for DNA recognition.

Keywords: DNA structure; transcription factors; indirect readout; protein–DNA interactions; gene regulation

Introduction

Sequence-dependent recognition of DNA by transcription factors and other DNA binding proteins is central to the regulation of gene regulatory processes. Direct physical interactions between DNA base pairs and protein residues have an important role in DNA sequence recognition, but the identities of non-contacted DNA bases can also markedly affect binding affinities.

The collective contributions of DNA sequence to the recognition process, referred to as indirect readout, are complicated to understand because they arise as a global consequence of the whole protein–DNA interaction.[1,2] Indirect readout is known to include sequence-dependent effects on DNA deformability as well as sequence-dependent effects on DNA conformation, local solvent and ion structure. However, the general guiding principles that underlie sequence-dependent protein–DNA recognition are far from entirely understood.

Analyses of X-ray crystal structures of protein–DNA complexes have provided many insights into the nature of protein–DNA recognition,[3,4] but these studies do not parse the extent to which the DNA structure adopted in the complex is induced by protein binding, or is instead intrinsic to the DNA sequence. To fully grasp how these effects separately contribute to protein–DNA recognition, an analysis of the DNA structure in both the free and bound state is necessary.

There are only a few examples of cognate DNA sequences with crystal structures reported in both

*Corresponding author. E-mail address: cathy.lawson@rutgers.edu.

† J.W.L., A.A.N., and S.C. contributed equally to this work.

Present addresses: J. Locasale, Department of Systems Biology, Harvard Medical School, Center for Life Science Rm. 417, 3 Blackfan Circle, Boston MA 02115, USA. E-mail address: jlocasal@bidmc.harvard.edu; A. Napoli, Clinical Connexion, 6 Terry Drive, Newtown, PA 18940, USA. E-mail address: andynap1@yahoo.com; S. Chen, Townsend and Townsend and Crew LLP, Two Embarcadero Center, San Francisco, CA 94111, USA. E-mail address: shengfeng_chen@yahoo.com.

Abbreviation used: MPD, methyl-2,4-pentanediol.

the free state and in the protein–DNA complex. These include the EcoRI endonuclease binding site,[5,6] the *trp* repressor operator region,[1,7,8] the papillomavirus E2 binding site,[9,10] runt-domain-containing transcription regulator binding sites,[11-14] and a nuclear factor kappa B binding site.[15,16] In each of these cases, unique insights into the specificity of the protein–DNA interaction were obtained by comparing bound and free DNA structures.

In this work, we carried out systematic crystallization screening of 50 DNA oligonucleotide duplexes with sequences derived from crystal structures of protein–DNA complexes (see Supplementary Data 1). By following up on positive hits from the screen, we obtained new crystal structures of cognate recognition sites for three distinct gene regulatory proteins: the methionine repressor (MetR), the purine repressor (PurR), and a zinc finger binding protein, Zif268. Each of these DNA binding proteins exhibits a different mode of DNA recognition. In each case, comparison of bound and free DNA permits identification of signatures (i.e., sequence-dependent conformational features) that must have a role in protein–DNA recognition. The new structures also permit comparison of the central ACGT tetrad that is shared by MetR, PurR, and E2 binding sites but is employed very differently in each of the three protein–DNA interactions. The DNA sequence immediately adjacent to ACGT appears to have a role in defining how ACGT is utilized in protein–DNA recognition. Our findings underscore the importance of sequence context in determining the contributions of DNA conformation and DNA deformability in protein–DNA interactions.

Results and Discussion

Recognition of the methionine repressor met box binding site

The MetR met box DNA dodecamer (free DNA site), d-(5′-GC**AGACGTCT**GC-3′)$_2$, was extracted from methionine repressor–DNA complex crystal structure DNA sequences d-(5′-TT**AGACGTCT**-3′)● d-(5′-**AGACGTCT**A-3′) (PDB ID 1cma[17]) and d-(5′-G**AGACGTCT**C-3′)$_2$ (PDB ID 1mjm[18], bound DNA site). The MetR dodecamer (free DNA site) crystallized in space group *P*1 with three duplexes in the asymmetric unit. The crystal structure was determined to 2.5 Å resolution (Table 1). The three independent dodecamers are essentially identical in structure and adopt a conformation close to ideal *B*-form DNA.

The MetR DNA dodecamer crystal lattice has a somewhat unusual packing arrangement. Two other DNA dodecamers (representing free binding sites

Table 1. Crystallographic data and refinement statistics

	Zif268 DNA decamer (FIN-A)	Zif268 DNA decamer (FIN-B)	MetR DNA dodecamer	purR DNA dodecamer
DNA duplex	5′-GCGTGGGCAC-3′ 3′-CGCACCCGTG-5′	5′-GCGTGGGACC-3′ 3′-CGCACCCTGG-5′	5′-GCAGACGTCTG-3′ 3′-GTCTGCAGACG-5′	5′-GCAAACGTTTGC-3′ 3′-CGTTTGCAAACG-5′
Space group	$P6_122$	$P6_122$	*P*1	$P3_121$
Crystal parameters				
a (Å)	38.26	38.48	39.99	39.64
b (Å)	38.26	38.48	40.37	39.64
c (Å)	78.17	78.23	40.27	98.59
α (°)	90	90	82.04	90
β (°)	90	90	69.13	90
γ (°)	120	120	81.49	120
X-ray diffraction				
Resolution limit (Å)	1.9	2.2	2.5	2.1
Temperature (K)	108	108	108	108
Radiation source	NSLS X12C	NSLS X12C	NSLS X26C	NSLS X8C; CuK_α
Wavelength (Å)	1.007	1.007	1.1004	1.0722; 1.5418
Reflections measured	41,269	10,824	33,594	54,956
Unique reflections	2905	1752	7790	5576
R_{merge}[a]	0.051	0.076	0.088	0.064
Completeness	0.973	0.868	0.893	0.946
Final model				
Asymmetric unit content	One duplex (50% occupancy)	One duplex (50% occupancy)	Three duplexes	One duplex
Number of atoms	434	425	1479	596
Solvent atoms	29	18	50	100
Bond lengths (Å)	0.017	0.010	0.014	0.002
Bond angles (°)	1.79	1.40	1.97	2.07
Torsion angles (°)	9.95	10.9	18.5	17.1
Average *B*-factor (Å^2)	21.1	23.1	62.7	41.6
R-factor[b]	0.215	0.230	0.235	.237
R_{free}[c]	0.251	0.280	0.283	.266

[a] $R_{merge} = \Sigma_h \Sigma_k | < I_k - I_{ki} > | / \Sigma_h \Sigma_i I_k$.
[b] $R = \Sigma | |F_o| - |F_c| | / \Sigma |F_o|$.
[c] $R_{free} = \Sigma_T | |F_o| - |F_c| | / \Sigma_T |F_o|$, where T is the test set comprised of 10% of observations excluded from refinement.

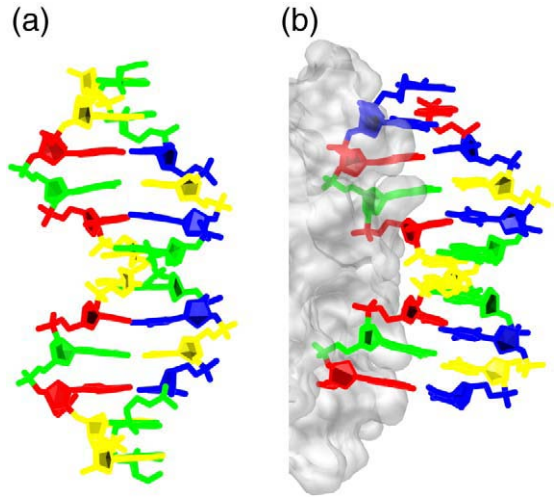

Fig. 1. Methionine repressor DNA binding site in free and protein-bound crystal structures. (a) The MetR dodecamer (5′-GCAGACGTCTGC-3′)$_2$ adopts a straight *B*-DNA structure (one of three independent copies is shown). (b) The protein-bound binding site is smoothly bent ~25° towards the major groove (PDB ID 1cma[17]). The DNA base color scheme is: red, adenosine; blue, thymidine; green, guanidine; and yellow, cytidine.

for papilloma virus E2 protein) also crystallize in space group *P*1 with similar lattice dimensions and three duplexes in the asymmetric unit;[10,19] the MetR DNA dodecamer crystallization condition, however, is unusual in its requirement for cobalt hexamine. Cobalt hexamine ions support the lattice structure by forming bridges between neighboring dodecamers. A similar arrangement has been reported in only one other *B*-form DNA crystal structure.[20]

The free MetR DNA dodecamer visually resembles the DNA in the protein-bound complexes, where two symmetry-related anti-parallel β strands of the MetR dimer are inserted into the major groove (Fig. 1). Relative to the free DNA, the bound DNA is bent smoothly towards the protein by ~25°, resulting in compression of the major groove and expansion of the minor groove. The root-mean-square (r.m.s.) deviation between the free and protein-bound DNA structures is ~2.0 Å (Supplementary Data 2 Fig. 1).

To compare the geometry of the free and bound structures, we considered four helical parameters: base pair, step, roll, and twist angles, and major and minor groove widths. Roll and twist angles are commonly used to characterize deformation from idealized DNA structures.[3] Major and minor groove widths provide additional measurements of helical geometry.

Plots of base pair step parameters across the common AGACGTCT sequence (Fig. 2) reveal structural patterns that appear to be shared by the free and bound DNA. Particularly striking are alternating regions of high and low values of major groove width and twist angle (Fig. 2a and d). The pattern is especially strong for the free DNA duplexes, where twist angles between consecutive base pair steps

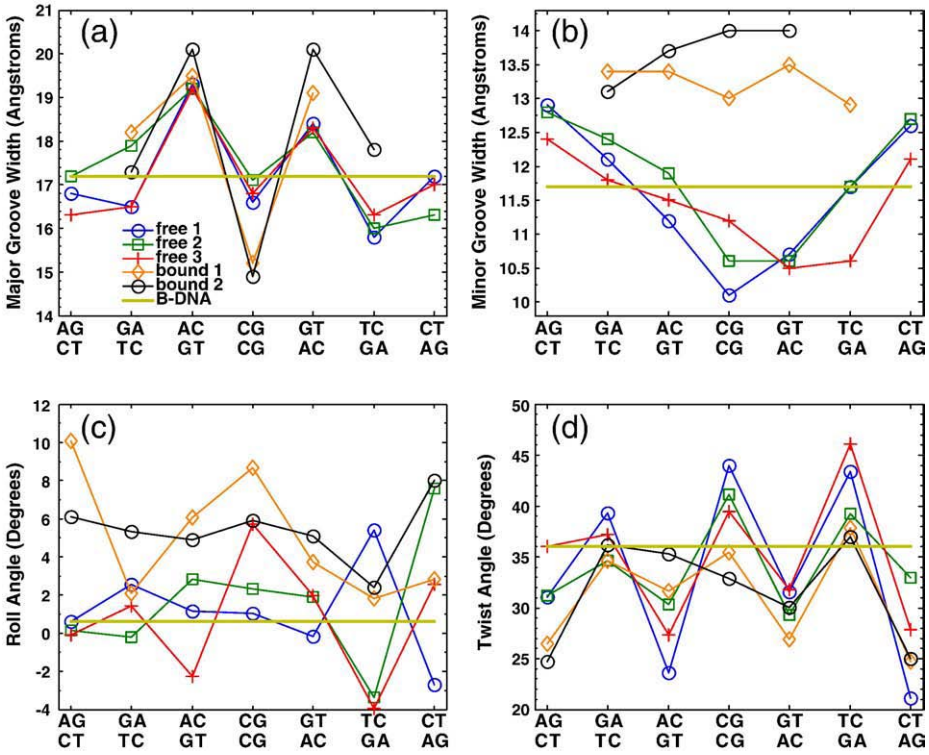

Fig. 2. Base pair step parameters of the MetR DNA binding site (5′-AGACGTCT-3′)$_2$ in free and bound crystal forms. (a) Major groove width; (b) minor groove width; (c) roll angle; (d) twist angle. Free 1, Free 2, and Free 3 are the three MetR DNA dodecamers in the crystal asymmetric unit; Bound 1 and Bound 2 are Met repressor/DNA complexes with PDB IDs 1cma[17] and 1mjm,[18] respectively.

can differ by up to 20°. Roll angle also displays some alternating behavior, though to a lesser extent (Fig. 2c). Minor groove width, on the other hand, appears uncorrelated between bound and free structures (Fig. 2b).

To further investigate the structural similarity of the free and bound MetR DNA binding sites, we calculated correlations for base pair step parameters across the common sequence between structure pairs. We first averaged the parameter values of the three free DNA duplexes (the parameter values do not vary significantly and considering each free structure individually does not alter the qualitative results). We then computed the correlation between the average free DNA and each protein-bound DNA structure (bound 1 = 1cma, bound 2 = 1mjm), as well as between the two bound structures, for each parameter (Fig. 3). The correlation values largely confirm trends apparent in Fig. 2. For every structure pair compared (average free *versus* bound 1, average free *versus* bound 2, bound 1 *versus* bound 2), major groove width patterns (Fig. 3a) and twist angle patterns (Fig. 2d) are highly correlated. Roll angle patterns are also correlated, though to a lesser extent (Fig. 2c). In contrast, there is no consistent trend for minor groove width pattern (Fig. 3b).

Together, these observations indicate that there is a consistent structural pattern shared by free and MetR bound DNA structures for the major groove width and twist angle base pair steps across the sequence AGACGTCT. The twist pattern follows the predicted trend based on statistical averages for base pair steps (average twist for AG = CT and AC = GT steps is ~32°, for GA = TC and CG steps, ~36°),[3] but the amplitude of the twist pattern is appreciably stronger, particularly in the free DNA structure. Our findings suggest that the MetR DNA binding site bears an intrinsic sequence-dependent conformational signature that is incorporated into the architecture of the protein-DNA.

The methionine repressor protein-DNA interaction involves an extensive network of direct and water-mediated hydrogen bonds to DNA bases and backbone phosphates, with two side-chains from each β strand contributing direct hydrogen-bonded contacts to the six central base pairs of the met box, GACGTC.[17,18] Mutation of the central CG step of the met box to a TA results in a ~50-fold decrease in binding affinity,[21] a strong indication of the sequence specificity of the interaction.

Contacts are also made by the methionine repressor to the sugar-phosphate backbone of the flanking AG = CT sequence of the met box, with phosphates being displaced by ~2 Å from the positions expected for ideal *B*-DNA. The positions of these phosphates appear to be essential in establishing the specificity of DNA recognition: any mutation of the central CTAG sequence within a double met box tandem operator (AGACGTCTAGACGTCT) decreases binding affinity by ~80-fold. On the basis of the protein–DNA complex structures alone, it was noted that these phosphate positions correspond to an alternating pattern of low and high twist angles that accompany an alternating contraction and expansion of the major groove.[18]

In the crystal structure of the free MetR DNA site, we observed alternating low and high twist angles that coincided with an alternating contraction and

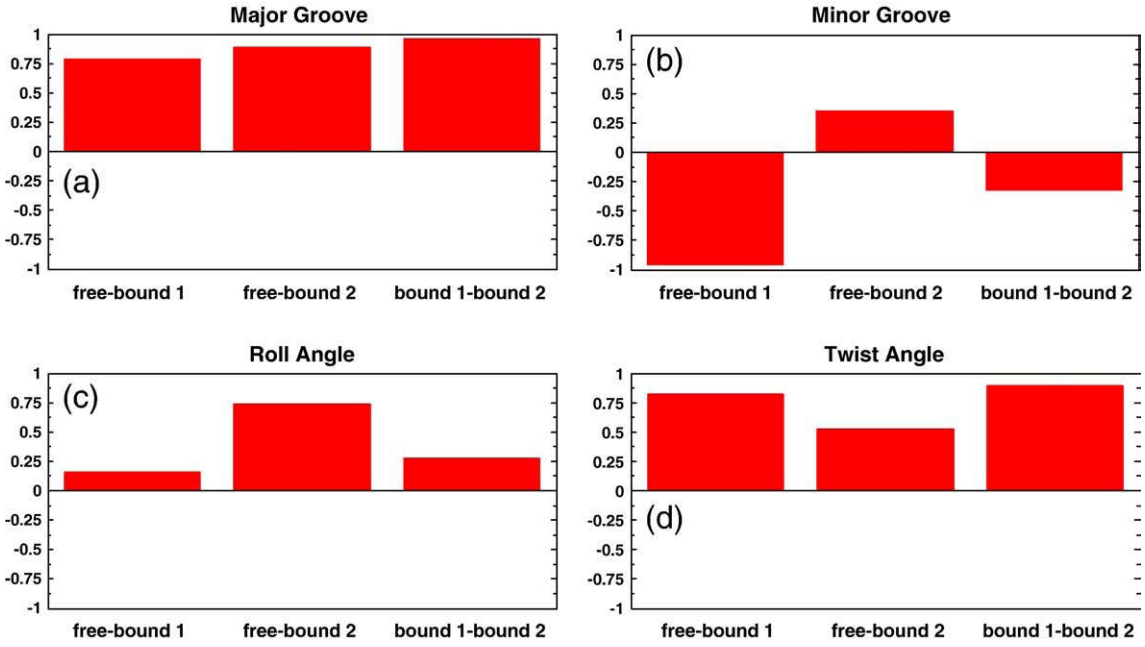

Fig. 3. Helical parameter correlations in MetR DNA structures. Correlation coefficients are shown for the helical parameter profiles for free DNA (free = averaged parameters for three independent dodecamers) *versus* bound DNA in 1cma (bound 1) and 1mjm (bound 2). (a) Major groove width; (b) minor groove width; (c) roll angle, (d) twist angle. Significant correlations are apparent in major groove width and twist angle.

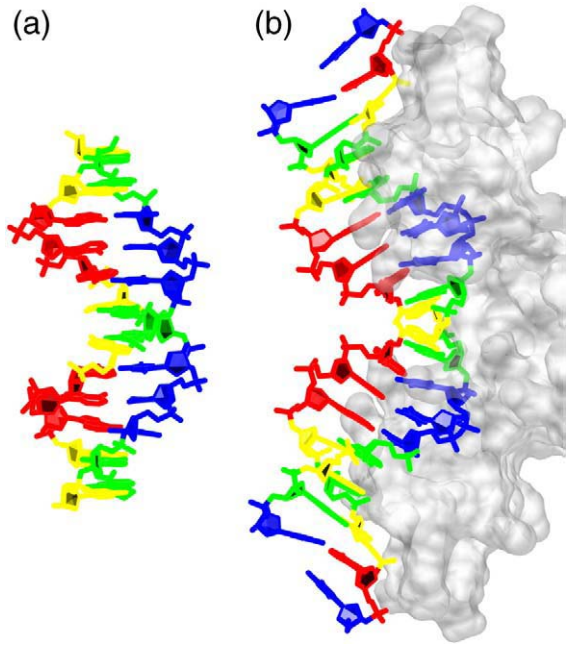

Fig. 4. Purine repressor DNA binding site in free and protein-bound crystal structures. (a) The PurR dodecamer (5′-GCAAACGTTTGC-3′)$_2$ adopts a straight *B*-DNA structure. (b) The protein-bound binding site[23] is bent ~50° towards the major groove at the central CG base pair (top right, PDB ID 1qpz). The DNA binding region of the purine repressor is shown as a semi-transparent surface. The color scheme is as described for Fig. 1.

expansion of the major groove, and the helical geometries of free and bound AGACGTCT met box sequences are strongly correlated. Our findings suggest that in addition to the backbone-mediated contacts to AG=CT, specificity is conferred through a partial lock-and-key fit between the GACGTC of the consensus sequence and the inserted β strands. The alternating twist pattern of the met box provides a surface poised to fit the β strand pair, requiring only the introduction of a small overall bend towards the major groove to achieve full protein–DNA complementarity.

Recognition of the purine repressor binding site

The free PurR DNA dodecamer, d-(5′-**GCAAACGTTTGC**-3′)$_2$, was extracted from the purine repressor–DNA complex crystal structure DNA sequence, d-(5′-TA**CGCAAACGTTTGC**GT-3′)$_2$ (PDB ID 1qpz).[22,23] The free PurR dodecamer crystallized in space group $P3_121$ with one complete dodecamer in the crystal asymmetric unit. The crystal structure was determined to 2.1 Å resolution (Table 1). This is a novel crystal form for dodecamer DNA with relatively high solvent content (volume per base pair = 1862 Å^3); the majority of DNA dodecamers crystallize in space group $P2_12_12_1$ (volume per base pair ~1390 Å^3).

The structure adopted by the free PurR dodecamer is also very close to that of ideal *B*-form DNA (Figs. 4a, and 5, blue squares). The average helical twist between successive pairs is 35.6°, and the

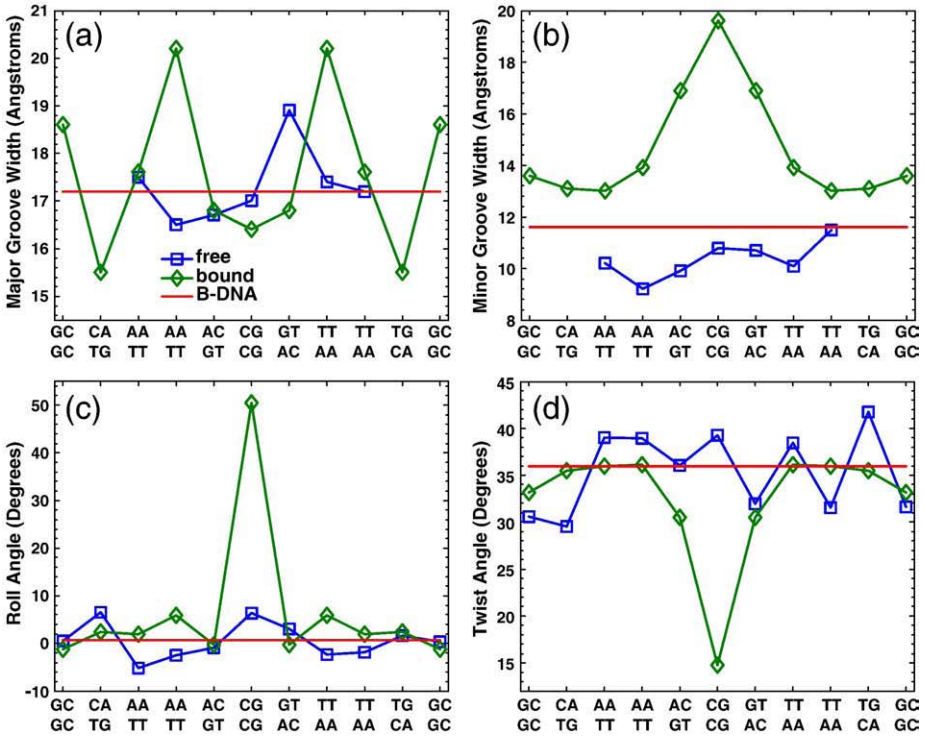

Fig. 5. Base pair step parameters of the PurR DNA binding site (5′-AAACGTTT-3′)$_2$ in free and bound crystal forms. Parameters are plotted *versus* base pair steps. (a) Major groove width; (b) minor groove width; (c) roll angle; (d) twist angle. Free, PurR DNA dodecamer; bound, purine repressor/DNA complex, PDB ID 1qpz.

average rise per base pair is 3.3°, resulting in a helix with 10.1 base pairs per turn. The CA=TG steps and the central CG step have slightly larger than average roll angles (~6°), in agreement with the observation that pyrimidine-purine base pair steps are more deformable than other base pair steps.[3] Both short A-tracts (AAA=TTT) feature a narrow minor groove (Fig. 5b) with a strong spine of hydration; the central A:T base pairs of the A-tracts have modestly high propeller twist (~−16°).

In the purine repressor dimer–DNA complex,[22,23] the bound DNA displays a sharp kink (Fig. 4b). This severe local deformation is accomplished by intercalation of a pair of leucine residues into the minor groove side of the central CG step and is characterized by a high roll angle for the central CG step (~50°), reduced twist angles for the central ACGT base pair steps, a narrowed major groove, and an enlarged minor groove (Fig. 5).

The central kink in the protein–DNA complex is the main conformational difference between the free and bound forms of the PurR binding site. Superimposition of the atoms of all 12 base pairs of the free and bound DNA structures yields a large r.m.s.d. (3.7 Å). However, when the bound structures is split at the central CG base pair step and the half-sites are superimposed onto the free structure, the r.m.s.d. is greatly reduced (1.3–1.4 Å). There is essentially no conformational change between free and bound DNA in the flanking GCAA=TTGC recognized by the purine repressor helix-turn-helix motifs (Supplementary Data 2 Fig. 2).

Three different crystal structures of a decamer duplex with the related sequence d-$(CCAACGTTGG)_2$ have been reported.[24,25] The central six base pairs of the free PurR dodecamer and the decamer structures (AACGTT) have very similar conformations (r.m.s.d. ~1.2 A, typical values for free *B*-DNA structures fitted to a fiber *B*-DNA model).[4] We conclude that in the free state, the central base pairs of the PurR DNA binding site prefer a conformation close to that of canonical *B*-DNA. The free PurR binding site structure is essentially equivalent to the bound PurR binding site structure with the localized kink removed.

ACGT motif within protein binding sites

The PurR and MetR binding sites share the common central tetrad ACGT. PurR bends ACGT dramatically towards the major groove via contacts in the minor groove, while MetR causes a milder deformation in the same direction via contacts in the major groove. ACGT also has a crucial role in the papillomavirus E2 protein interaction with its binding site (CGACCG**ACGT**CGGTCG).[9,10] E2 bends the central ACGT towards the minor grove, with negative roll values at all three central base pair steps. E2 makes no direct contact with ACGT, yet the sequence is essential for recognition. ACGT also forms the center of recognition sites for leucine zipper proteins GCN4 (ATG**ACGT**CAT) and *max* (TC**ACGT**GA), where the protein-bound DNA has essentially a canonical *B*-DNA form.[26,27] These findings suggest that although ACGT tends to adopt the *B*-DNA conformation in the absence of protein contact, it is a highly deformable sequence that can be utilized in dramatically different ways.

The extended sequence context of ACGT appears to have an important role in how it is utilized in protein–DNA recognition. ACGT is embedded in three different octamer sequences with free and bound structure representatives, PurR: AA**ACGT**TT, MetR: AG**ACGT**CT, and E2: CG**ACGT**CG. The relatively rigid short flanking A-tracts in the PurR binding site likely aid localization of the protein-induced kink to the central four base pairs, serving as a structural buffer between the central hinge-helices-DNA interaction and flanking helix-turn-helix-DNA interactions. In contrast, flanking pyrimidine-purine CG base pair steps of the E2 binding site extend flexibility, allowing protein-induced deformation to be distributed over the entire octamer. The flanking AG base pair steps of the free MetR binding site extend an alternating twist pattern inherent in the central ACGT (compare Figs. 2d and 5d), forming a distinctive recognition surface that is retained with overall deformation in the protein–DNA complex.

Recognition of Zif268 zinc finger binding sites

The free decamers FIN-A, d-(5′-**GCGTGGGCAC**-3′)●d-(5′-**GTGCCCACGC**-3′) and FIN-B, d-(5′-**GCGTGGGACC**-3′)●d-(5′-**GGTCCCACGC**-3′) contain binding sites for the zinc finger-containing protein Zif268. The sequences were extracted from crystal structures of Zif268 bound to DNA duplexes d-(5′-A**GCGTGGGCAC**-3′)●d-(5′-T**GTGCCCACGC**-3′), (PDB ID 1a1h) and d-(5′-A**GCGTGGGACC**-3′)●d-(5′-T**GGTCCCACGC**-3′), (PDB ID 1a1f), respectively.[28]

Both free decamers crystallized in space group $P6_122$ with a half decamer in the crystal asymmetric unit. By difference Fourier analysis, the crystals were shown to have discrete disorder about the crystallographic 2-fold axis (see Materials and Methods). Full decamer duplex structures with half occupancy were obtained for FIN-A to 1.9 Å resolution and for FIN-B to 2.2 Å resolution (Table 1). Both decamer structures are similar to canonical *A*-form DNA[29] and to each other (r.m.s.d. 1.1–1.3 Å). The crystal packing of the free decamers is typical for *A*-DNA and potentially stabilizes the *A*-form.[30]

Free and bound Zif268 DNA binding sites are shown in Fig. 6 (see also Supplementary Data 2 Figs. 3 and 4). The FIN-A and FIN-B DNA decamers adopt *A*-DNA conformation, though with a narrower minor groove than is typical for *A*-DNA. In the Zif268–DNA interaction, three consecutive zinc fingers wrap around the ten base pair consensus site, each inserting one α-helix into the major groove.[28] The bound DNA structures are largely *B*-form although, as reported earlier, they have some *A*-like character, including a wider minor groove than is typical for *B*-DNA.

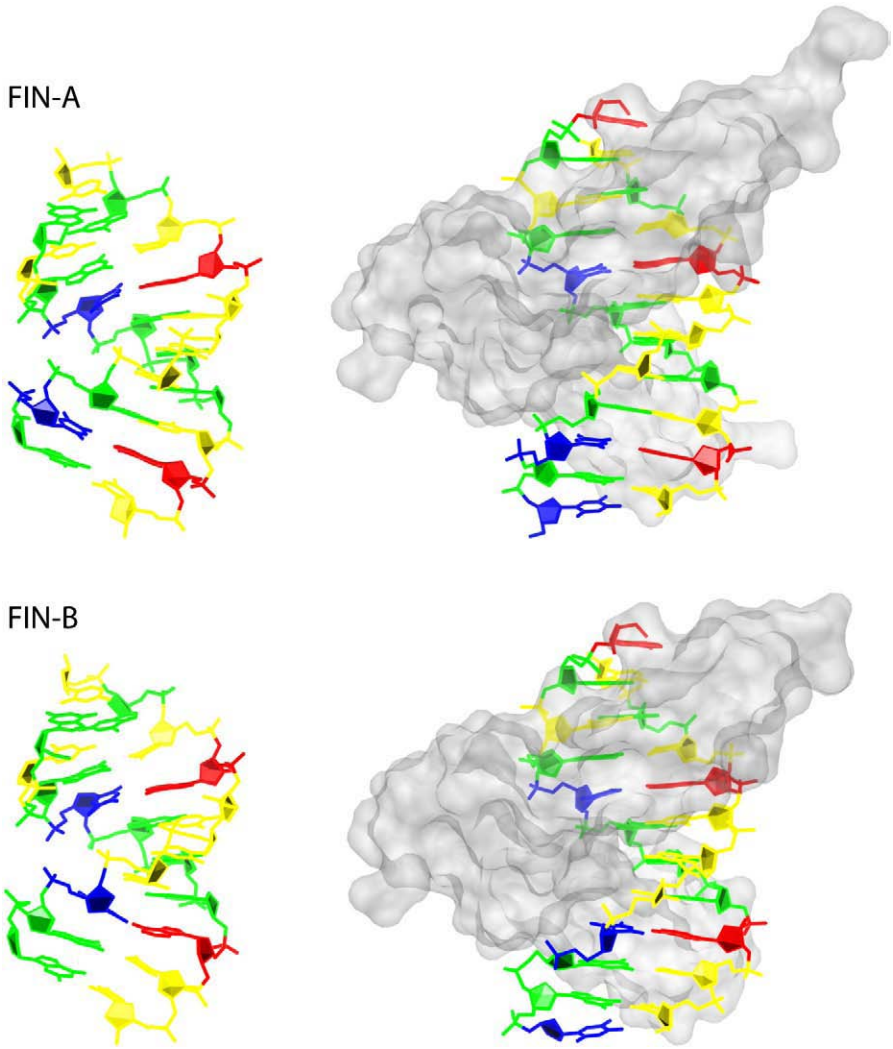

Fig. 6. Structures of Zif268 zinc-finger DNA binding sites in free and protein-bound crystal forms. The free (5′-GCGTGGGCAC-3′)●(5′-GTGCCCACGC-3′) Zif268 zinc-finger binding site decamer (FIN-A, top left) and the free (5′-GCGTGGGACC-3′)●(5′-GGTCCCACGC-3′) Zif268 zinc-finger binding site decamer (FIN-B, bottom left) adopt *A*-DNA structures, while the corresponding protein-bound binding sites (top right, PDB ID 1a1h; bottom right, PDB ID 1a1f)[28] adopt *B*-DNA conformations with *A*-like character. The color scheme is as described for Fig. 1.

The *A*-*B* transition has been investigated in a number of DNA structures, and may be biologically significant, at least as an indication of intrinsic deformability.[8,13,31-34] Two parameters used to describe where a structure lies in the *A*-*B* continuum are *x*-displacement (displacement of a base pair away from the helix axis in the direction of the base pair short axis),[13,34,35] and Zp (mean *z*-coordinate of the backbone phosphorus atoms with respect to individual base pair dimer reference frames).[31,36] By both of these criteria, the free and bound Zif268 DNA binding sites appear as intermediates, with the free sites closer to canonical *A*-DNA and the bound sites closer to canonical *B*-DNA, although the spread is more convincing for base-pair geometry (x-displacement, Fig. 7a) than for backbone conformation (Zp, Fig. 7b).

The structural transition from *B* to *A* DNA involves an increase in roll angle from 1.5° to 12.0° and a decrease in slide from +0.5 Å to −1.4 Å.[36] A plot of roll *versus* slide for individual base pairs (Fig. 7c) shows that both free and bound Zif268 DNA binding sites fall between canonical *A*-DNA and canonical *B*-DNA conformations. The free Zif268 DNA binding sites have *A*-DNA-like slide values but intermediate roll values, while the bound Zif268 DNA binding sites have slide and roll values intermediate between those of canonical *B*-DNA and canonical *A*-DNA.

The dodecamer d-$(CATGGGCCCATG)_2$ has been characterized as an *A*/*B* intermediate.[37,38] The free dodecamer shares with the Zif268 binding sites the internal sequence TGGG; parameters for dodecamer TGGG are included in Figs. 7 and 8 for comparison. As determined by the backbone conformation, the dodecamer TGGG is very close to canonical *A*-DNA and the free Zif268 binding sites (Fig. 7b). But by base pair geometry criteria, the dodecamer TGGG is clearly an intermediate between the bound and free Zif268 binding sites,

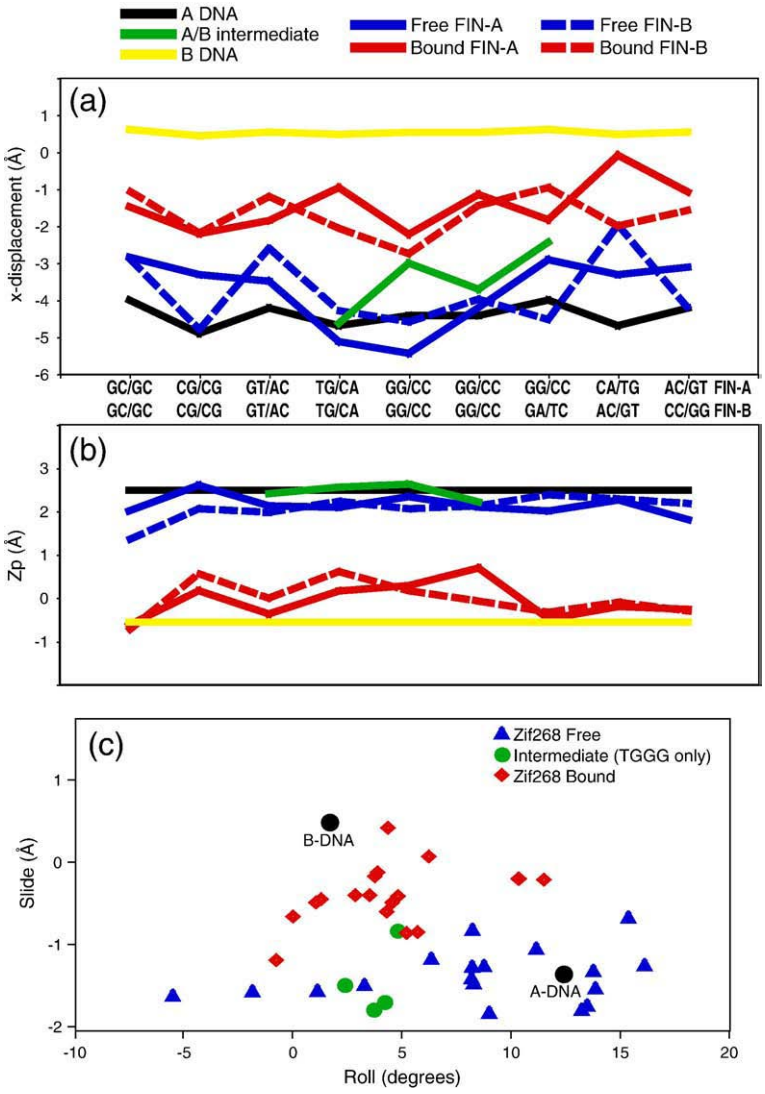

Fig. 7. Base pair step parameters of Zif268 zinc finger free and bound binding sites. (a) X-displacement; (b) Zp; (c) roll *versus* slide. Values are also plotted for the internal TGGG of the d-(CATGGGCCCATG)$_2$ The DNA structure is the intermediate *A*/*B* form,[38] and values for canonical *A*-DNA and *B*-DNA are indicated.[53]

both in terms of *x*-displacement (Fig. 7a) and roll *versus* slide (Fig. 7c).

A/*B* hybrid binding sites

An investigation of bound *versus* free binding sites for another GC-rich DNA sequence that tends to adopt *A*-like conformation, the runt-domain binding site (TGCGGTC), has been carried out by the Shakked group.[11,13] Even though details of the protein–DNA interaction differ (IgG-like runt domains bind to the major groove of their seven base pair recognition site via a β-hairpin), the results from the runt-domain system strongly parallel ours for the Zif268 system. Free runt-domain DNA binding sites were observed in two distinct *A*/*B* intermediate conformations: in nonamer duplexes, TGCGGTC adopts *A*-like structures,[13] while in a dodecamer duplex, the same sequence is closer to *B*-DNA conformation, albeit with some *A*-like character.[11] TGCGGTC in runt-domain bound DNA duplexes is also close to the *B*-DNA conformation, but with somewhat more *A*-like character than the free dodecamer.

Using average roll *versus* slide as an indicator, runt-domain binding sites and Zif268 binding sites cluster into two distinct *A*/*B* intermediate groups within the *A*-*B* continuum (Fig. 8). One free site (the runt-domain dodecamer) plus all four bound sites from the two systems form a cluster with conformations closer to *B*-DNA; all other free sites form a second cluster with conformations closer to *A*-DNA. The *A*/*B* intermediate d-(CATGGGCCCATG)$_2$ dodecamer[38] and *A*/*B* intermediate d-(GGCGCC)$_2$ derivative structures[34] form a third (uncircled) cluster at the lower-left corner of the roll *versus* slide plot, with positions intermediate between the two runt/Zif268 binding site clusters.

The Zif268 and runt-domain binding sites both feature central G-tracts, which are commonly found in crystals of *A*-DNA, but are rarely found in crystals of *B*-DNA.[30] The triplet GGG, in particular, has not been crystallized in canonical *B*-DNA conformation except within protein–DNA complexes. Both of these DNA-binding sequences have base pair step trimers that have been defined as *A*-philic (GGG for Zif268 and GGT for runt-domain) and both also lack trimers defined as *B*-philic.[39] We conclude that the

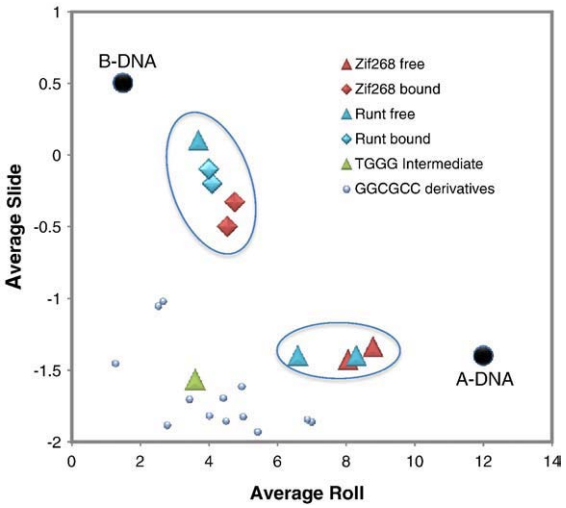

Fig. 8. Comparison of free and bound intermediate *A/B* DNA structures. Average roll (°) *versus* average slide (Å) are plotted per duplex for free (red triangles) and bound (red diamonds) Zif268 binding sites, and free (cyan triangles) and bound (cyan diamond) binding sites for Runt. Two clusters formed by the Zif268 and runt binding site structures are indicated with ovals. For reference, average roll *versus* average slide are plotted for the TGGG-containing *A/B* intermediate (green triangle, PDB ID 1dc0),[38] and *A/B* intermediate structures of d-(GGCGCC) derivatives[34] (gray dots, PDB ids 1ih2, 1ih3, 1ih4, 1ih5, 1ih6).

particular DNA sequences in these two binding sites strongly influence local DNA conformation and deformability, enabling hybrid *A/B* conformations closer to *B*-DNA or closer to *A*-DNA, depending on the specific environmental condition. We predict that either binding site can make the transition between *A/B* intermediate forms via a small energy barrier, and that either conformation can exist in solution.

Conclusions

Using a structural genomics approach, we targeted cognate DNA binding sites of known protein–DNA complexes and obtained four new crystal structures of free DNA binding sites, corresponding to three distinct protein-binding sites. By this method, we have significantly increased the number of DNA recognition motifs of proteins for which both bound and free structures are available.

For each of the three recognition motifs examined, the DNA is recognized in a fundamentally different manner, yet in each case we observed significant structural correlations between free and bound structures. The results strongly underscore the importance of DNA sequence context in protein–DNA recognition.

In the case of the MetR binding site, a distinctive geometric pattern with alternating twist angles, roll angles, and major groove width is shared by the free and bound DNA structures. Strong correlations in the helical geometry of free and bound DNA structures are reported here for the first time and provide evidence for a structural code of DNA recognition in the form of a constrained geometry in the major groove.

In the case of the PurR binding site, the free DNA duplex crystallized in a nearly ideal *B* form, nearly identical with the bound DNA duplex but without the prominent central kink present in the protein–DNA complex. The observation that the free DNA structure exhibits no kink is consistent with the current model that indirect readout occurs primarily through protein-induced bending. Short flanking A-tracts provide conformational rigidity to help isolate the central deformation.

In the case of the Zif268 binding sites, the free structures reveal an intrinsic tendency to adopt hybrid *A/B* DNA structure, facilitating recognition and binding of the target transcription factor, which also recognizes an *A/B* intermediate conformation. The Zif268 binding sites, like those for runt,[13] demonstrate that the intrinsic tendency of certain DNA sequences to adopt *A/B* hybrid conformations can have an important role in protein–DNA recognition. Deformation in the *A-B* continuum represents an alternative strategy to deformation by bending for optimization of complementarity between protein and DNA binding surfaces.

Materials and Methods

Selection of target DNA sequences

A total of 50 DNA decamer and dodecamer duplexes were designed for crystallization trials on the basis of sequence recognition motifs represented in one or more known protein–DNA crystal structures. The DNA duplexes were screened for crystallization under standard conditions. The sequences are given in Supplementary Data Table 1, and further descriptions of the design strategy, screening efforts, and results are provided in the accompanying text.

DNA synthesis and annealing

DNA oligonucleotides purified by C-18 dual reverse-phase HPLC were purchased from Integrated DNA Technologies, Inc (IDT). DNA duplexes were prepared in 5 mM sodium cacodylate (pH 6.5), 10 mM $MgCl_2$, 5 mM EDTA to yield final concentrations of 0.5–1.0 mM duplex. The duplexes were annealed by heating at 90 °C for 5 min, and then cooled slowly to 25 °C over a period of 12 – 15 h.

Data collection, structure determination and analysis

All crystals were prepared by the hanging-drop, vapor-diffusion method, were mounted on nylon loops and flash-cooled to 100 K. Except where noted below, X-ray diffraction data were collected at the Brookhaven National Laboratory National Synchrotron Light Source (NSLS). Intensity data were processed using DENZO and SCALEPACK,[40] and model refinement was performed

with CNS[41] using maximum likelihood refinement, bulk solvent correction, anisotropic temperature factor scaling, and the DNA parameter restraints as described.[42] Manual model building was performed using O[43] or XtalView.[44] Solvent atoms consistently observed in omit maps within 3.4 Å of nitrogen or oxygen atoms were retained in final models. The crystal parameters, diffraction and model refinement statistics are summarized in Table 1. DNA helical parameters were calculated using 3DNA.[45] Uncertainties in helical parameters resulting from coordinate error are expected to be small relative to the differences discussed in the main text.[45]

MetR DNA dodecamer

Crystals of dodecamer (5′-GCAGACGTCTGC-3′)$_2$ were grown in 10% (w/v) methyl-2,4-pentanediol (MPD), 40 mM sodium cacodylate (pH 5.5), 20 mM cobalt hexamine, 20 mM $MgCl_2$. The reservoir contained 500 μl of 35–55% MPD. Before data collection, a single crystal was equilibrated against a reservoir of 45% MPD. Native X-ray diffraction data were collected to 2.5 Å resolution at NSLS beamline X26C.

The structure was solved by molecular replacement using the program EPMR.[46] The coordinates of NDB entry BD0001[10] (5′-ACCGACGTCGGT-3′)$_2$ were used as the search model. Placement of three duplexes with corrected sequence in the *P*1 asymmetric unit yielded a starting *R*-factor of 35 %. After rigid body and simulated annealing refinement, inspection of $2F_o - F_c$ and $F_o - F_c$ electron density maps revealed the presence of three ordered cobalt hexamine molecules, one in the major groove of each duplex. After completion of refinement with CNS, TLS parameters[47] were refined for each independent duplex as a group using Refmac.[48]

For comparison of base pair step parameters between free and bound MetR binding sites, the elements of the correlation matrix *C* were calculated as:

$$C_{ij} = \left(N_{\text{tot}}\sqrt{\langle x_i x_i\rangle\langle x_j x_j\rangle}\right)^{-1} \sum_{\{\text{base pair steps}\}} (x_i - \langle x_i\rangle)(x_j - \langle x_j\rangle)$$

C_{ij} gives a quantitative measure of the degree of cross-correlation between the different structures (x_i and x_j are parameter values for structures *i* and *j*, brackets denote an average across the different base pairs, and N_{tot} is the total number of base pair steps that are considered).

PurR DNA dodecamer

Crystals of dodecamer (5′-GCAAACGTTTGC-3′)$_2$ were grown in 1 mM spermine tetrahydrochloride, 30% MPD, 40 mM magnesium acetate, 50 mM sodium cacodylate (pH 6.0). Before data collection, a single crystal was transferred to 1 mM spermine tetrahydrochloride, 45% MPD, 40 mM magnesium acetate, 50 mM sodium cacodylate (pH 6.0). Native X-ray diffraction data were collected to 2.1 Å resolution at NSLS beamline X8C. To improve completeness at low resolution, the synchroton data were merged with data measured from a second native crystal using a Quantum 1 CCD detector with mirror-focused CuK_α radiation from a rotating anode source.

Multiwavelength anomalous dispersion data were collected to 2.1 Å resolution from a single crystal of the bromine derivative (5′-GCAAACGT(5-Br-U)TGC-3′)$_2$ at NSLS X8C. Data were measured at three incident radiation wavelengths corresponding to peak, inflection and remote wavelengths of the bromine K-edge. The two Br atoms were located by inspection of an anomalous difference Patterson map, and the relative positions of the atoms and absolute hand of lattice symmetry were found by difference Fourier methods. Phase refinement was carried out using MLPHARE, with the inflection dataset designated as native.[49]

Solvent-flattening and histogram-matching density modification procedures were applied to native amplitudes and multiwavelength anomalous dispersion-derived phases using the program DM.[50] The resulting electron density map revealed clear density for a duplex of 12 separate base pairs in the $P3_121$ asymmetric unit. A fiber *B*-DNA dodecamer model generated using 3DNA[45] was manually fit into the density. Initial structure refinement used a combined amplitude and density-modified phase maximum likelihood target.

Zif268 DNA decamers (FIN-A, FIN-B)

Crystals of the FIN-A decamer (5′-GCGTGGGCAC-3′)●(5′-GTGCCCACGC-3′) were grown in 10% (w/v) MPD, 40 mM sodium cacodylate (pH 5.5), 20 mM cobalt hexamine, 12 mM NaCl, 80 mM KCl. Crystals of the FIN-B decamer (5′-GCGTGGGACC-3′)●(5′-GGTCCCACGC-3′) were grown in 10% MPD, 40 mM sodium cacodylate (pH 5.5), 20 mM cobalt hexamine, 80 mM NaCl, 20 mM $MgCl_2$. Crystals of the FIN-A decamer bromine derivative (5′-GCG(5-Br-U)GGGCAC-3′)●(5′-G(5-Br-U)GCCCACGC-3′) were grown in either 10% MPD, 40 mM sodium cacodylate (pH 5.5), 20 mM cobalt hexamine, 20 mM or 10% MPD, 40 mM Mes (pH 6.0), 20 mM cobalt hexamine, 20 mM $MgCl_2$. In all cases, the reservoir was composed of 35% MPD. Before data collection, crystals were equilibrated against a reservoir of 45% MPD. Native datasets were collected at NSLS X12C from a single crystal of decamer A to 1.9 Å resolution and from a single crystal of decamer B to 2.2 Å resolution. FIN-A decamer bromine derivative data were collected on an R-axis IV detector with Yale mirror-focused CuK_α Rigaku RU100 rotating anode source (statistics not shown). Merging statistics for the native and derivative datasets were consistent with the space group assignment $P6_122$ (or $P6_522$).

Plausible molecular replacement solutions were found for both FIN-A and FIN-B decamer crystals with the program AMoRe,[51] using the coordinates of the symmetric decamer (5′-GACCGCGGTC-3′)$_2$ (NDB ID AD0001[52]) as the search model. In each case, the molecular 2-fold symmetry axis of the search model was found to coincide with a crystallographic 2-fold symmetry axis of a $P6_122$ lattice. Difference Fourier maps revealed discrete disorder for both bromine atoms of the FIN-A derivative about the putative crystallographic 2-fold, confirming the space group assignment. For refinement, initial coordinate sets were constructed as full DNA duplexes with half occupancy. The starting models yielded *R*-values of 32% and 35% for FIN-A and FIN-B, respectively. After an initial round of rigid body refinement, alternating rounds of simulated annealing with torsion angle dynamics, energy minimization, *B*-factor refinement of domains, and manual model adjustments were performed. At later stages, *B*-factors were refined for base pair groups, and finally for individual atoms. The final structures have pseudo 2-fold symmetry with overlapping base pairs in the crystal lattice sharing the same planes, and overlapping base-pair steps sharing similar roll and slide values. Any error in estimation of helical parameters resulting from ambiguity in atomic positions would be expected to be small.

Protein Data Bank accession codes

The atomic coordinates and structure factors file has been deposited in the PDB with accession codes 1HQ7 for PurR dodecamer (NDB ID BD0047), 2B1D for MetR dodecamer (NDB ID BD0090), 2B1B for FIN-A decamer (NDB ID AD0058), and 2B1C for FIN-B decamer (NDB ID AD0059).

Acknowledgements

We are grateful for the assistance of Jin Wu, Irina Persikova, Satish Bodige, and Suzanne Richman in carrying out crystallization trials. We thank Leon Flaks, Annie Heroux, and Anand Saxena for their assistance with data collection at National Synchrotron Light Source beamlines X8C, X12C, and X26C, which are supported by the Department of Energy and the National Institutes of Health (NIH). This study was supported by NIH award GM21589 to C.L.L. and H.M.B. A.A.N. was an NIH trainee (GM08319).

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.01.007

References

1. Otwinowski, Z., Schevitz, R. W., Zhang, R. -G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q. *et al.* (1988). Crystal structure of *trp* repressor/operator complex at atomic resolution. *Nature*, **335**, 321–329.
2. Lawson, C. L. & Berman, H. M. (2008). Indirect readout of DNA sequence by proteins. In *Protein Nucleic Acid Interactions* (Rice, P. A. & Correll, C. C., eds), pp. 66–86, Royal Society of Chemistry, Cambridge, UK.
3. 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.
4. Jones, S., van Heyningen, P., Berman, H. M. & Thornton, J. M. (1999). Protein-DNA interactions: a structural analysis. *J. Mol. Biol.* **287**, 877–896.
5. Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1981). Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl Acad. Sci. USA*, **78**, 2179–2183.
6. Kim, Y., Grable, J. C., Love, R., Greene, P. J. & Rosenberg, J. M. (1990). Refinement of Eco RI endonuclease crystal structure: a revised protein chain tracing. *Science*, **249**, 1307–1309.
7. Lawson, C. L. & Carey, J. (1993). Tandem binding in crystals of a *trp* repressor/operator half-site complex. *Nature*, **366**, 178–182.
8. Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A. & Sigler, P. B. (1994). Determinants of repressor operator recognition from the structure of the *trp* operator binding site. *Nature*, **368**, 469–473.
9. 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.
10. Rozenberg, H., Rabinovich, D., Frolow, F., Hegde, R. S. & Shakked, Z. (1998). Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets. *Proc. Natl Acad. Sci. USA*, **95**, 15194–15199.
11. Bartfeld, D., Shimon, L., Couture, G. C., Rabinovich, D., Frolow, F., Levanon, D, Shakked, Z. *et al.* (2002). DNA recognition by the RUNX1 transcription factor is mediated by an allosteric transition in the RUNT domain and by DNA bending. *Structure*, **10**, 1395–1407.
12. Bravo, J., Li, Z., Speck, N. A. & Warren, A. J. (2001). The leukemia-associated AML1 (Runx1)–CBF beta complex functions as a DNA-induced molecular clamp. *Nat. Struct. Biol.* **8**, 371–378.
13. Kitayner, M., Rozenberg, H., Rabinovich, D. & Shakked, Z. (2005). Structures of the DNA-binding site of Runt-domain transcription regulators. *Acta Crystallogr. D*, **61**, 236–246.
14. Tahirov, T. H., Inoue-Bungo, T., Morii, H., Fujikawa, A., Sasaki, M., Kimura, K, Ogata, K. *et al.* (2001). Structural analyses of DNA recognition by the AML1/Runx-1 Runt domain and its allosteric control by CBFbeta. *Cell*, **104**, 755–767.
15. Chen, Y. -Q., Ghosh, S. & Ghosh, G. (1998). A novel DNA recognition mode by the NF-KB p 65 homodimer. *Nat. Struct. Biol.* **5**, 67–73.
16. Huang, D. B., Phelps, C. B., Fusco, A. J. & Ghosh, G. (2005). Crystal structure of a free kappa B DNA: Insights into DNA recognition by transcription factor NF-kappa B. *J. Mol. Biol.* **346**, 147–160.
17. Somers, W. S. & Phillips, S. E. V. (1992). Crystal structure of the met repressor-operator complex at 2.8 Å resolution reveals DNA recognition by β-strands. *Nature*, **359**, 387–393.
18. Garvie, C. W. & Phillips, S. E. V. (2000). Direct and indirect readout in mutant Met repressor-operator complexes. *Struct. Fold. Des.* **8**, 905–914.
19. Hizver, J., Rozenberg, H., Frolow, F., Rabinovich, D. & Shakked, Z. (2001). DNA bending by an adenine–thymine tract and its role in gene regulation. *Proc. Natl Acad. Sci. USA*, **98**, 8490–8495.
20. Gao, Y. G., Robinson, H., Sanishvili, R., Joachimiak, A. & Wang, A. H. J. (1999). Structure and recognition of sheared tandem G×A base pairs associated with human centromere DNA sequence at atomic resolution. *Biochemistry*, **38**, 16452–16460.
21. Wild, C. M., McNally, T., Phillips, S. E. & Stockley, P. G. (1996). Effects of systematic variation of the minimal *Escherichia coli* met consensus operator site: in vivo and in vitro met repressor binding. *Mol. Microbiol.* **21**, 1125–1135.
22. Glasfeld, A., Koehler, A. N., Schumacher, M. A. & Brennan, R. G. (1999). The role of lysine 55 in determining the specificity of the purine repressor for its operators through minor groove interactions. *J. Mol. Biol.* **291**, 347–361.
23. Schumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. (1994). Crystal structure of LacI family member, PurR, bound to DNA: minor groove binding by alpha helices. *Science*, **266**, 763–770.
24. Chiu, T. K. & Dickerson, R. E. (2000). 1 Å crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium. *J. Mol. Biol.* **301**, 915–945.

25. Privé, G. G., Yanagi, K. & Dickerson, R. E. (1991). Structure of the *B*-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G. *J. Mol. Biol.* **217**, 177–199.
26. Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B. & Burley, S. K. (1993). Recognition by max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature*, **363**, 38–45.
27. Keller, W., Konig, P. & Richmond, T. J. (1995). Crystal structure of a bZIP/DNA complex at 2.2 Å: determinants of DNA specific recognition. *J. Mol. Biol.* **254**, 657–667.
28. Elrod-Erickson, M., Benson, T. E. & Pabo, C. O. (1998). High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure*, **6**, 451–464.
29. Arnott, S. (1999). Polynucleotide secondary structures: an historical perspective. In *Oxford Handbook of Nucleic Acid Structure* (Neidle, S., ed), pp. 1–38, Oxford University Press, Oxford, UK.
30. Wahl, M. C. & Sundaralingam, M. (1997). Crystal structures of A-DNA duplexes. *Biopolymers*, **44**, 45–63.
31. Lu, X. -J., Shakked, Z. & Olson, W. K. (2000). A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* **300**, 819–840.
32. Malinina, L., Fernandez, L. G., Huynh-Dinh, T. & Subirana, J. A. (1999). Structure of the d(CGCCCGCG-GGCG) dodecamer: a kinked A-DNA molecule showing some B-DNA features. *J. Mol. Biol.* **285**, 1679–1690.
33. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.
34. Vargason, J. M., Henderson, K. & Ho, P. S. (2001). A crystallographic map of the transition from B-DNA to A-DNA. *Proc. Natl Acad. Sci. USA*, **98**, 7265–7270.
35. Dickerson, R. E. & Ng, H. L. (2001). DNA structure from A to B. *Proc. Natl Acad. Sci. USA*, **98**, 6986–6988.
36. Calladine, C. R. & Drew, H. R. (1997). *Understanding DNA: The Molecule and How it Works*, 2nd edit. Academic Press, San Diego, CA.
37. Ng, H. L. & Dickerson, R. E. (2002). Mediation of the A/B-DNA helix transition by G-tracts in the crystal structure of duplex CATGGGCCCATG. *Nucl. Acids Res.* **30**, 4061–4067.
38. Ng, H. L., Kopka, M. L. & Dickerson, R. E. (2000). The structure of a stable intermediate in the A ↔ B DNA helix transition. *Proc. Natl Acad. Sci. USA*, **97**, 2035–2039.
39. Tolstorukov, M. Y., Ivanov, V. I., Malenkov, G. G., Jernigan, R. L. & Zhurkin, V. B. (2001). Sequence-dependent B to A transition in DNA evaluated with dimeric and trimeric scales. *Biophys. J.* **81**, 3409–3421.
40. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
41. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Warren, G. L. *et al.* (1998). Crystallographic and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
42. Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. (1996). New parameters for the refinement of nucleic acid containing structures. *Acta Crystallogr. D*, **52**, 57–64.
43. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.
44. McRee, D. E. (1999). XtalView/Xfit – a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165.
45. Lu, X. J. & Olson, W. K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31**, 5108–5121.
46. Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D*, **55**, 484–491.
47. Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D*, **57**, 122–133.
48. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D*, **53**, 240–255.
49. Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In *Proceedings of the CCP4 Study Weekend* (Wolf, W., Evans, P. R. & Leslie, A. G. W., eds), pp. 80–86, SERC Daresbury Laboratory, Warrington, UK.
50. Cowtan, K. (1994). DM: an automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newsletter Protein Crystallogr.* **31**, 34–38.
51. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A*, **50**, 157–163.
52. Finley, J. B. & Luo, M. (1998). X-ray crystal structures of half the human papilloma virus E2 binding site: d(GACCGCGCTC). *Nucleic Acids Res.* **26**, 5719–5727.
53. Chandrasekaran, R. & Arnott, S. (1989). The structures of DNA and RNA helices in oriented fibers. In *Landolt-Börnstein Numerical Data and Functional Relationships in Science and Technology, Group VII/1b, Nucleic Acids* (Saenger, W., ed), Springer-Verlag, Berlin.