Published in final edited form as:
J Mol Biol. 2014 October 23; 426(21): 3553–3568. doi:10.1016/j.jmb.2014.08.009.

DNA-recognition by a σ^{54} transcriptional activator from *Aquifex aeolicus*

Natasha K. Vidangos[a], **Johanna Heideker**[a], **Artem Lyubimov**[a], **Meindert Lamers**[a], **Yixin Huo**[c], **Jeffrey G. Pelton**[b], **Jimmy Ton**[a], **Jay Gralla**[c], **James Berger**[a], and **David E. Wemmer**[a,b,*]

[a]Department of Chemistry, MC-1460, University of California, Berkeley, CA 94720, USA

[b]Physical Biosciences Division, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA

[c]Department of Chemistry and Biochemistry and Molecular Biology Institute, University of California, Box 951569, Los Angeles, California 90095, USA

Abstract

Transcription initiation by bacterial σ^{54}-polymerase requires the action of a transcriptional activator protein. Activators bind sequence-specifically upstream of the transcription initiation site via a DNA-binding domain. The structurally characterized DNA-binding domains from activators all belong to the Factor for Inversion Stimulation (Fis) family of helix-turn-helix DNA-binding proteins. We report here structures of the free and DNA-bound forms of the DNA-binding domain of NtrC4 (4DBD) from *Aquifex aeolicus*, a member of the NtrC family of σ^{54} activators. Two NtrC4 binding sites were identified upstream (−145 and −85 base pairs) from the start of the *lpxC* gene, which is responsible for the first committed step in Lipid A biosynthesis. This is the first experimental evidence for σ^{54} regulation in *lpxC* expression. 4DBD was crystallized both without DNA and in complex with the −145 binding site. The structures, together with biochemical data, indicate that NtrC4 binds to DNA in a manner that is similar to that of its close homologue, Fis. The greater sequence specificity for the binding of 4DBD relative to Fis seems to arise from a larger number of base specific contacts contributing to affinity than for Fis.

INTRODUCTION

The efficient, effective regulation of gene expression is a vital process in an organism's ability to survive in a changing environment. In prokaryotes, gene expression is most highly

*To whom correspondence should be addressed: Tel: (510) 666-2683; dewemmer@berkeley.edu.

Accession numbers
The coordinates for the structures have been deposited in the Protein Data Bank with the following IDs: unbound structure: 3E7L; DNA-bound structure: 4FTH.

SUPPLEMENTARY DATA
Supplementary data are available at the Journal of Molecular Biology online: Supplementary Figures 1, 2, 3.

regulated at the transcriptional level. All transcription requires the RNA polymerase (RNAP) holoenzyme, which includes a catalytic core of five subunits and an additional, variable sigma factor that conveys promoter specificity.[1] Different sigma factors are used in the cell to transcribe different sets of genes.[2] This allows coordinated regulation, an important component of the fine-tuning of gene expression. The sigma factors fall into two families, σ^{70} and σ^{54}. The σ^{70} family is represented in all bacteria, generally with many variants differing in size and regulatory mechanism in any given organism.[2] The σ^{70} holopolymerase is inherently competent to bind target promoters, and spontaneously open DNA and initiate transcription without the action of any other proteins. By contrast, the σ^{54} form of polymerase, present in many, but not all bacteria, is also competent to bind promoters, but it is inherently unable to melt the template DNA promoter and initiate transcription. For initiation to occur, σ^{54} requires interaction with a transcriptional activator.[2,3,4] The transcriptional activators usually consist of three domains: a sensory domain, a central AAA+ ATPase domain, and a C-terminal helix-turn-helix DNA-binding domain (DBD).[5] The DNA-binding domain often binds to two high-affinity cooperative binding sites located ~100 basepairs upstream of the σ^{54} binding site.[6] Upon activation of the sensory domain, the activator ATPase domain acts on σ^{54}, and induces a conformational change that enables σ^{54} to melt the DNA and initiate transcription.[5] σ^{54} regulation allows for a wide dynamic range of transcription, and rapid temporal response to changing conditions. It has been found to control a variety of processes, including nitrogen starvation response in *Xylella fastidiosa*[7] and *E. coli*,[8] metabolism of xylenes and toluenes in *Pseudomonas putida*,[9] and the expression of virulence factors that support flagellar motility and pilus-mediated attachment in the bacterial pathogen *Pseudomonas aeruginosa*.[10]

A great deal of structural and biochemical work has been devoted to understanding the mechanisms for regulation and ATPase activity of σ^{54} transcriptional activators.[5,11] However, there is little known about activators' DNA binding domains, particularly bound to their target DNA. The only structures of activator DBDs reported and described in detail to date are from NtrC,[12] although mutations used to stabilize the protein eliminated DNA binding capability, and from ZraR,[13] both from *Salmonella typhimurium*. These have helix-turn-helix folds, butressed by a third helix, as seen in many other bacterial DNA binding proteins.[14] In both NtrC and Fis there is a fourth helix that is important for dimerizing the DBDs, and holding the recognition helices with a spacing close to that of two successive major grooves of DNA. No structures of activator DBDs have been reported in complex with DNA. Here we present the structure of the DBD of NtrC4 (4DBD) from the extreme thermophile *Aquifex aeolicus*, both alone and in complex with a DNA oligomer containing the sequence from an upstream region of the *lpxC* gene, whose product, the LpxC deacetylase, catalzes the first committed step in Lipid A biosynthesis.[15] We reported the overall fold of the free 4DBD previously, together with the NtrC4 receiver and ATPase domains,[11] but its structure was not discussed in detail.

The DBDs of NtrC proteins are close homologues and evolutionary ancestors of the versatile DNA-binding and bending protein, Factor for Inversion Stimulation (Fis).[16] During exponential growth, Fis is one of the most abundant proteins in enteric bacteria, and is involved in a number of processes ranging from site-specific recombination, integration-

excision reactions of phages, and negative transcriptional regulation, to cell cycle timing in chromosome replication.[17] While Fis can bind nonspecific sequences with high affinity (nanomolar dissociation constants), certain sequences can form especially “stable” high-affinity complexes that are not disrupted by the addition of competitor DNA.[18,19] The ability to bind both these preferred sequences as well as many others at varying affinities is a common characteristic of histone-like DNA-binding proteins such as Integration Host Factor (IHF) and HU. However, within this class of histone-like proteins, each protein’s requirement for consensus sequences and sequence-derived DNA flexibility varies. For example IHF shows a clear preference for binding consensus sequences, but HU does not.[20] Fis appears to be an intermediate example, and its ability to bind non-specifically with high affinity makes it challenging to identify preferences for binding particular sequences. Using different prediction methods, there have been conflicting predictions of 6,000 and 68,000 Fis-binding sites in the *E.coli* genome.[18,19]

Several structures of Fis alone were solved,[21, 22, 23, 24] and there is a multitude of biochemical data describing its binding to a variety of nonspecific DNA-binding sequences, and its potency in bending DNA by as much as 90°.[21, 25, 26]

The similarity in sequence and structure of Fis and NtrC family DBDs opens the question of how these two closely related protein families perform very different functions with different requirements for binding specificity. For NtrC, DNA-bending has been observed with electron microscopy,[27] but no high resolution structural or biochemical studies have been performed to characterize the degree of bending. Additionally, there is no evidence that NtrC proteins bind DNA nonspecifically with significant affinity or perform any of the additional chromatin-like DNA-restructuring functions of Fis. Therefore, the structural details that lead to the difference in behavior of these two close homologues will aid our general understanding of how and why proteins bind and bend DNA.

Understanding of how NtrC- and Fis-like DNA-binding proteins bind and distort DNA, are of intrinsic interest, but could also have application in affecting σ^{54}-controlled virulence factors or disruption of Fis-modulated celllular functions. In spite of the interest in understanding the details of NtrC-like protein complexes with DNA since the time the initial structures were solved, there has been only one NMR study exploring backbone structure and residues involved in the DNA-binding of the σ^{54}-activator NifA,[28] but no structure of a complex. The first high-resolution structures of Fis complexes were solved recently by Stella et al. [29] The structures of Fis bound to a high-affinity site, and bound to a number of binding site variants, provided characterization of protein-DNA interactions and the mode of DNA recognition. A model was presented in which Fis searches the DNA for regions with narrow minor grooves, binds, and then compresses the minor groove further, creating the characteristic FIS-induced bend. The minimal reliance on specific base contacts and the large number of DNA backbone contacts provide some explanation as to the ability of Fis to bind without significant sequence specificity.

The structures of the free and bound DNA-binding domain of NtrC4 allow a detailed comparison with Fis to understand the similarities and differences that modulate binding specificity and DNA distortion. These structures, and comparisons with other structures in

the family, provide new insights into to how σ^{54} transcriptional activators bind DNA and ultimately activate transcription.

RESULTS

Crystal structure of the free 4DBD

The crystal structure of the 4DBD (encompassing residues 365–442 of the full length NtrC4 protein) was determined to resolution of 2.25 Å (Table 1). The 4DBD crystallized with two dimers in the asymmetric unit, with a total of four zinc atoms bridging the protein dimers. These ions are not required for structure or stability (verified in NMR and unfolding studies, data not shown), but were convenient for phasing. For the discussion of structures we use the amino acid numbering in the actual construct studied (full length 365 = residue 1, etc.). In keeping with helix-turn-helix nomenclature, each monomer consists of four helices, A – D, with the D helix representing the “recognition helix” that is typically inserted into the major groove of DNA (see Figure 1A). The A and B helices of 4DBD form an antiparallel four-helix bundle with the corresponding helices in the adjacent monomer. As already observed for the DNA-binding domain of NtrC,[12] the A helices (Leu19-Lys24) do not overlap significantly but pack against the long B helix (Lys29 – Tyr47) through hydrophobic sidechain interactions. Helices C (Leu51 – Ile58) and D (Leu62 – Leu72) form a classic helix-turn-helix DNA binding motif near the C- terminus of the molecule. The HTH motif is extended by a tail consisting of residues Leu72 to Ser79. Dimer formation buries an area of 2160 $Å^2$, corresponding to 30% of the total surface area of 4DBD. The same values are observed for the DNA binding domain of NtrC (30%) and Fis (30%).

In the 4DBD structure no density could be observed for residues Met1 – Trp16 and only weak density was visible for Leu77-Ser79, indicating that the N- and C-termini are disordered. To further analyze this disorder, NMR relaxation was used to assess backbone dynamics. The backbone resonances were assigned for 4DBD, and T1, T2, and NOE values were measured for each residue. NOEs from residues 1–15 and 77–79 were either negative or small (under 0.5), compared to an average value of 0.9 in the core of the protein. T1 and T2 relaxation experiments showed slow relaxation in residues 3–14 and 77–79. Moreover the ^{15}N NOE becomes negative for residues 2–5 and 78–79 (Supplemental Figure 1). These data confirm that the first 15 residues of 4DBD are very flexible in solution. This region corresponds to the linker region between the DBD and ATPase domains, a region with low sequence conservation among transcriptional activators and is variable in length (Supplemental Figure 2). Flexibility in this region is consistent with a requirement for flexibility to accommodate the reorganization of the ATPase domains when the protein goes from inactive dimers to active hexamers.

The overall fold of the 4DBD is similar to Fis and NtrC-DBD[12] (Fig. 1b). Superposing the three-helix bundle (residues Lys29 – Leu72) with the corresponding residues in Fis and NtrC using the program Superpose[30] gives root mean square deviations of 0.65 Å and 2.1 Å, respectively.

Dimerization occurs mainly through hydrophobic interactions between the A and B helices. Contacts between the end of the B helix and the loop between helix C and D may also

contribute to dimer stabilization. Close intra-monomer contacts are observed between the sidechains of Leu22 and Leu19 in helix A and Ile38 of helix B. Close inter-monomer contacts are made between the sidechains of Leu19, Leu23 and Ile38. The B helices of each monomer cross at a 40° angle such that the phenyl rings of residue Phe35 of each monomer stack upon each other in a parallel displaced manner.

Determination of the DNA-binding site for NtrC4

In order to study a complex of 4DBD with cognate DNA, its specific binding sites were identified. Potential upstream activation sequences (UASs) were selected based on the presence of σ^{54} binding sites in the promoter region. Using the software PATSER[31], the *A. aeolicus* genome was analyzed using a σ^{54} consensus sequence mrNrYTGGCACG-N4-TTGCWNNw (capital letters indicate high conservation, lowercase indicates weak conservation, and the nucleic acid codes are as follows: m = A/C, r = G/A, n = A/C/G/T, y = T/C, w = A/T).[6] Candidate σ^{54}-binding sites were identified in the promoters of six *A. aeolicus* genes: *nirB*, *fhp*, *glnB*, *dhsU*, *Aq_087*, and *lpxC*. The first five sites had been identified as σ^{54}-binding sites previously,[32] but *lpxC* did not appear in the previous analysis. Since σ^{54}-activators often bind ~100 bp upstream of the σ^{54} consensus site,[6] DNA regions extending 250 base pairs upstream of these five promoter regions were amplified by PCR and subjected to gel-shift assays with full-length NtrC4 (4FL) and the NtrC4 DNA-binding domain alone (4DBD). Gel shifts indicate that NtrC4 binds only to the *lpxC* promoter with sufficient affinity to cause a gel shift at a concentration of 50 nM (Supplemental Figure 2).

To determine the specific binding site, DNase I footprinting was performed on the 4DBD-*lpxC* upstream region. Two areas of significant protection were visible centered around 145 and 85 bp upstream of the transcription start site (Supplementary Figure 3) at a protein concentration of 10 nM. To further confirm these binding sites, DNA oligomers containing the two sites, and a series of mutant sites, were purchased with a 5′-fluorescein tag and affinity binding assays were performed. Resulting K_d values are summarized in Figure 4.

Structure determination of the 4DBD-variant_1 complex

Since flexible regions can prevent crystallization, a new 4DBD construct (374–442), which removed the first 9 flexible residues, was used for complex formation. The 4DBD complex crystallized in $P2_1$ symmetry and the structure was solved at 3.0 Å resolution in complex with a 22mer DNA that contained a two-base-pair mismatched overhang (Forward 5′-**AC**TTGCAAATTTGCAAATGCAT-3′; Reverse 5′-**TG**ATGCATTTGCAAATTTGCAA-3′). Due to the DNA-mismatches in the overhangs, the first (5′)-adenine base of the forward strand and (5′)-thymine base of the reverse strand flipped out of the helix, and the resulting one-base-pair G/C overhang self-annealed, forming a continuous double helix through the crystal lattice with 21 base pairs. Although a variety of sequences of different lengths and varying overhangs crystallized, these samples gave poorer resolution (>4Å). Since the flipped-out bases appear to be disordered in the structure and are not clearly visible in the density, the numbering convention used in this paper will describe only the 21 nucleotides observed (Figure 3a).

4DBD-DNA complex: protein-DNA interactions

As in the free protein, the N-terminal residues preceding Arg16 (residues 10–15 in the short construct) are disordered. Omitting this region, the free and bound forms of 4DBD are very similar in structure, with an overall RMSD of 0.57 Å. The most notable differences are the orientations of the sidechains of R67 and Y66. Compared to the free protein, Arg67 extends outward towards the DNA bases in the complex structure, and Y66 rotates by ~ 90° to contact the DNA backbone.

The structure of the bound protein, and expansions showing some of the protein-DNA interactions are presented in Figures 1 and 2. As is expected for helix-turn-helix proteins,[14] the majority of protein-DNA interactions occur in the D helices. Since the protein is a symmetric dimer, most interactions with one half-site binding sequence are repeated in the other half-site. For simplicity, interactions in only one end of the helix will be described below, except in the cases where asymmetry exists in the protein-DNA interactions.

To enable simple comparisons between the different DNA-binding half sites of 4DBD and Fis, the following base-numbering convention will be used throughout this discussion. Bases in the native binding site used for crystallography (“variant_1”) are labeled from −8 to +12, centered around the palindromic center of the 14-base pair “core” binding site, and a parallel numbering convention is presented for Fis (Figure 3). Both bases in a given Watson-Crick base pair will bear the name of the same position shown in Figure 3 (from −8 to +12) and the strand (F, forward, or R, reverse) will be specified separately. For example, −7T(F) refers to the second base on the forward strand, and is base-paired with −7A(R), the last base on the reverse strand.

The primary sequence-specific contacts are made by Arg67 and Ser63. The two sidechain NH_2 groups of Arg67 form hydrogen bonds with the O6 atoms of +4G(F) and +5G(R), bridging the two DNA strands. The hydroxyl oxygen of Ser63 accepts a hydrogen bond from the amino group of +5C(F), which is buttressed by a hydrogen bond to the sidechain carboxyl group of Asp61 (See Fig 2d). Asn64 also forms a hydrogen bond with the N7 group of +4G(F), forming a contact that selects for a purine (Figure 2d). In addition to these hydrogen bonds, Asn64 is held in place by van der Waals contacts to Arg67 and +4G phosphate backbone. This combination of a hydrogen bond network and van der Waals contacts produce a structurally well-defined protein-DNA binding interface at the D helices.

There are also a number of nonspecific contacts with the DNA phosphate backbone. To orient the recognition helices in the major groove, Lys68 hydrogen bonds with the phosphate group on the +3T(F) backbone. Lys70 hydrogen bonds with the phosphate group of +7T(R). Further stabilization of the recognition helix comes from a hydrogen bond formed between the backbone amide of Asp61 and the phosphate of G12(F).

The only other nonspecific contacts outside of the 4DBD D helix are a hydrogen-bonding ring between residues in both protein monomers (distinguished as A and B) that form the following array of hydrogen bonds with a thymine phosphate: −3T(R) to 32Lys(A) to 36Glu(B) to 68Lys(B) to −3T(R) (Figure 2c). All of these interactions are repeated in the symmetrical monomeric subunit. The only notable difference between the two subunits is

the conformation of the Tyr66 residue, which forms a hydrogen bond with the phosphodiester O on one side of the monomer +8T(R), but bonds directly with the final free phosphate on the other monomer +1C(F), due to the asymmetrical presence of the DNA nick on the 5′F end (Fig 2a, b).

4DBD-DNA complex: DNA conformation

The structure of the DNA in the complex shows only slight deviations from B-form DNA. It is important to note that although the symmetrical protein monomers overlay closely, the fact that the actual binding site is not at the center of the oligonucleotide that crystallized causes slight structural differences at the two ends. On the 5′-end of the forward strand, the break in the DNA occurs only two base pairs away from the −5G and −4C base pairs (the "GC motif"). On the 3′-end of the forward strand, there are an additional 9 base pairs following the GC motif. This asymmetry is also reflected in features of DNA distortion that are discussed below.

The DNA-binding site has an A-tract separating the two GC motifs that participate in specific protein contacts, a sequence motif that is relatively common as a target for helix-turn-helix DBDs. It has been shown that regions of four or more adenosine residues in a row form structures that are relatively inflexible. A variety of different studies have shown that periodic A-tracts can form stably-curved DNA, although the precise basis for these features are still debated.[33] The A-tract sequences, coupled with YR steps, are one of the most common elements in sequences that undergo distortion by proteins.[34] The DNA in the 4DBD complex exhibits an overall bend of ~9°, with the majority of bending occurring at base pairs+5 to +7, which are toward the 3′ end of the A-tract.

Most of the DNA in the 4DBD-DNA structure is close to B-form duplex, however, there are two base pairs, located at the +8 and +12 positions, which show significant shear. The +8 base pair is in a short A-tract following the second GC pair, the beginning of which is the region of maximal DNA bending. The +12 base pair is the last base pair before reaching the DNA overhangs, and its distortions are likely the product of crystal contacts that cause the overhangs to self-anneal.

It has been consistently observed that pyrimidine(Y)-purine(R) base pair steps are more deformable than YY, RR, or RY steps.[35] The "TGCA" motif that repeats in the binding site for 4DBD provides two points of flexibility (T-G and C-A), enabling the compression of the minor groove and expansion of the major groove. Plots of DNA roll, bend, and major/minor groove widths at positions along the DNA are presented in Figure 5.

Biochemical analysis of the DNA-binding site of NtrC4

To probe the relative importance of different base pairs in the DNA-binding site, gel-shift assays were performed on 26-bp fluorescein-labeled DNA duplexes with changes introduced at specific base pairs (Figure 4). In three of the sites (variant_1, variant_3, variant_4), the curves did not fit a two-state model (i.e., the data indicate no measurable binding at low protein concentrations, followed by an abrupt increase in binding at intermediate concentrations). As a result, these K_d values are included only as qualitative estimates, and have an asterisk (*) to highlight this fact. The native complex was found to exhibit a K_d of

11 ± 4 nM (variant_2; Figure 4b). As expected, a single mutation of the specificly contacted G-C base pairs (located at positions +/−4 and +/−5) to A-T reduced binding affinity 10-to-15-fold (variant_3 and variant_4 (both non-two state), variant_5, and variant_6). A disruption of the A tracts that are outside of the GC sequences (position +/−7) caused a nearly 20-fold reduction in binding affinity (variant_7), suggesting the importance of the peripheral A-tracts and DNA flexibility in these regions. In contrast, the insertion of CG bases in the middle of the central A-tract (positions −1 and +1, variant_8) has a less dramatic effect on binding, reducing affinity only 5-fold.

DISCUSSION

The free 4DBD has wider recognition-helix-spacing than Fis

It is informative to compare the structures of the free Fis and 4DBD. One important difference between these structures is that the recognition helices in 4DBD have a larger separation between helix axes, ~24 Å, than in Fis, ~20 Å (measured by a least-distance function for helices in UCSF-Chimera). Without protein or DNA distortion, the 4DBD D-helices are able to fit into sequential major grooves of ideal B-form DNA (gap ~ 21 Å), whereas Fis would not be able to fit into neighboring major grooves without compression of the straddled minor groove.[19] It is not surprising, therefore, that 4DBD produces a much subtler bend in the DNA than Fis. Interestingly, the DNA-binding domain of another homolog, NtrC from *Salmonella typhimurium*, has an inter-D-helix spacing of only ~14 Å.[12] Although no structure of a complex with DNA is available for it, there is evidence for significant induced DNA bending by NtrC.[27] Whereas Fis is understood to exhibit DNA remodeling functions in some circumstances, the functional consequences of NtrC bending of DNA are not fully understood.

Fis-DNA and 4DBD-DNA complexes exhibit similar specific contacts

The interactions between protein and DNA in 4DBD and Fis complexes exhibit a number of key similarities. In both systems an arginine residue (Arg67 in 4DBD, and Arg85 in Fis) is the dominant specific contact to a guanine base. Both also have an asparagine residue near the N7 group of a purine. In the case of Fis, Asn84 forms a hydrogen bond with +4G. Mutation of +4G to a C removes this hydrogen bond, but does not have a dramatic effect on binding affinity. However, mutating +4G to a T causes a steric clash that reduces overall binding affinity ~2500-fold.[17] Therefore, it appears that the Fis asparagine residue serves to prevent binding at sites with a T residue at ±4. The same may be true for 4DBD, which has an asparagine residue (Asn64) that appears to be close enough to form three hydrogen bonds: one to the N7 purine group of +4G (2.9Å N-N distance), one to the phosphate backbone of +4G (3.0 Å N-O), and one to the hydroxyl group of Ser63 (3.8Å N—O). Higher resolution data are necessary to precisely define the interactions of this residue and assess contributions of hydrogen bonds vs. van der Waals interactions. However, it is apparent that if the ±4 residue were a thymine, the 5′ methyl group would produce a steric clash with Asn64 that would require a rearrangement of the residue that would likely reduce binding affinity. This may be a second cause, together with the loss of a specific contact to the +4G base, for reduction of the 4DBD binding affinity ~10 fold in a +4G->+4T mutation (Figure 4, variant_5).

In Fis, direct interaction of arginine residues with guanines causes kinks with high roll angles (13°, 9°) at the GA steps. Although 4DBD bends the DNA much less than Fis, the roll angle profile for 4DBD is quite similar (see Figure 5). In 4DBD, high roll values occur at the YR steps surrounding the G basepairs (−6T to −5G and −4C to −3A). The bend angles were also calculated with Curves+[36] and are compared in Figure 5. The 3′F end of the DNA shows a gradual curve at the peripheral A-tract of approximately 2° per base pair, in a pattern very similar to Fis.

Fis and 4DBD distort the DNA minor grooves similarly

Stella et al. describe that the DNA in the Fis-DNA complex has expanded minor grooves opposite the major grooves where the interaction takes place, and a compressed central minor groove. They also find that the major groove widths are more or less constant, allowing the minor groove compression to lead to a global bend in the DNA. The 4DBD complex profile appears quite similar (Figure 5c). Differences between the DNA-bending profiles in the FIS and 4DBD complexes appear to result from two features: 4DBD protein binding near the end of the DNA strand, and the different placement of specific contacts in the binding site.

As can be seen in Figure 3a, the DNA-binding domain binds near the 5′ forward end of the DNA, likely causing distortions in the 5′ forward end due to crystal packing contacts. The central A-tracts are fairly straight, but the CA base steps at −4 to −3 and +5 to +6 are sites of the most deformability and are overtwisted with high roll angles, generating a gradual bend, predominantly in the A-tract at positions +6 to +8. Disruption of the peripheral A-tracts that border these overtwisted bases (including bases +6 to +8) produces reduction of binding affinity by ~20-fold when both peripheral A-tracts are disrupted (Figure 4, variant_7). This combination of DNA base pair conformations produces a 10° overall global bend. Given the convergence of protein-binding and crystal packing contacts that take place near the −8 end, we expect that the DNA distortions occurring at +1 to +8 positions are the most illustrative of native binding. Due to the symmetry of the protein and palindromic binding site, we would therefore predict that if the complex were crystallized with fewer inter-DNA crystal contacts and with a longer DNA-binding site at the −8 position, the bend seen in the +5 to +8 region would be repeated in the −5 to −8 region, producing a slightly more pronounced global bend, but still less than that induced by Fis.

The relative placement of specific contacts in each complex is also notably different. In the 4DBD complex, specific contacts are clustered in two areas, whereas the FIS complex has four specific contacts that are spread further apart (Figure 3a). This leads to a shorter bent region in the 4DBD complex relative to the FIS complex (Figure 5c). Whereas 4DBD has 6 base pairs between the two specifically-contacted GC base pairs, Fis has 13 base pairs between the specifically contacted guanines. Also, in 4DBD, the Arg67 sidechains point inward, toward the center of the binding site, to contact the GC motifs. In Fis, the Arg85 sidechains point outward, almost a 90° rotation from the Arg67 position in 4DBD (Figure 3b). The other specific contact in 4DBD, Ser63, has no equivalent residue in the Fis complex. However, there is close structural overlap between the following nonspecific contacts: Asn64, Asp61, and Tyr66 from 4DBD are in similar positions as Thr87, Asn84,

and Arg89, respectively. As a result, the overall binding profile shows very similar nonspecific contacts, but with sidechains that "reach outward" to make specific contacts in the Fis complex versus contacts that "reach inward" to make specific contacts in the 4DBD complex.

NtrC4 and Fis appear to bind DNA by different mechanisms, but distort DNA similarly

Stella et al. have proposed that Fis binds DNA by scanning sequences for narrow minor grooves, binding sequences that have a loosely defined consensus sequence, and then inducing a bend in the DNA to generate a tight fit. It appears unlikely that 4DBD follows this same mechanism. The separation between the D helices in 4DBD is significantly wider than that of Fis, leaving enough space to accommodate a B-form minor groove without steric clashes. The observed small bend, and the modest sensitivity of binding affinity to a DNA site with a disrupted central A-tract (Figure 4c, variant_8) suggest that bendability of the central minor groove is not the determining factor of binding. Instead, the complex's reliance on a greater number of residues involved in specific contacts in 4DBD (eight contacts to bases, mediated by six residues) vs. Fis (eight contacts to bases, mediated by two residues) likely increases binding specificity. Fis also contains more nonspecific contacts to the backbone (twelve) than 4DBD (eight). One of the most important residues for nonspecific interactions in Fis, Lys90,[37] has no structural equivalent in 4DBD.

In NtrC residues identified as important for DNA binding through mutagenesis (R456, N457, R461)[38] are in the recogniton helix, and are predicted to point into the groove, the difference in spacing of the helices[39] leaves unclear what bending is induced by NtrC and which residues are responsible. Localized bending may affect the cooperativity of binding activators at adjacent sites, and hence may affect potency of activator sites. Although the NtrC DNA-binding domain has been shown to bind some "governor sequences" nonspecifically to prevent DNA looping and inhibit transcriptional activation, it has only been observed at high concentrations and with low affinity (K_d > 650 nM).[40] There have been no other indications in the literature that NtrC-like proteins can bind DNA nonspecifically at concentrations lower than 650 nM. The mostly likely role of DNA bending in activator binding to enhancers is on cooperativity of binding. Enhancers generally have two binding sites for dimers, which can vary in spacing. Both binding affinity and cooperativity in occupation of these sites affect the enhancer strength. Cooperativity probably depends upon contacts between dimers at neighboring sites, and is enhanced when the proteins are in their activated state in which ATPase domains also interact. DNA bending may be needed to allow inter-dimer contacts, particularly when unactivated. Interactions between DNA binding domains and ATPase domains are poorly understood, but one study with electron microscopy suggests a change in ordering of DNA binding domains that depends on the nucleotide state of attached ATPase domains assembled into active rings for NtrC[41]. DNA bending in the region between the enhancer and the promoter start site is also important, but is mediated by IHF[42] and is probably not related to bending at the enhancer sites.

The structural differences described above, which suggest greater sequence specificity and reduced ability to bend DNA for 4DBD relative to Fis, are consistent with the different

functions of the two proteins. The specificity of binding of NtrC activators is a critial aspect of their function. In contrast, the myriad of functions exercised by Fis require a dynamic binding mechanism that enables binding to many sequences with varying extents of DNA bending.

NtrC4 may regulate *lpxC* transcription

The *lpxC* gene is involved in the biosynthesis of Lipid A, the hydrophobic anchor of lipopolysaccharide, which forms the outer monolayer of the outer membrane of most gram-negative bacteria.[15] Since this outer membrane in gram-negative bacteria is implicated in numerous forms of antibiotic resistance, it has sparked interest as an antibiotic target.[43] The finding that NtrC4 binds upstream of the *lpxC* promoter appears to be the first experimental evidence suggesting σ^{54} control of genes related to cell-surface lipid metabolism, although infomatics analysis has suggested a relationship between σ^{54} and the bacterial cell exterior[44]. In our previous work on the regulation of NtrC4, we found that it can assemble to an active oligomer even in the absence of phosphorylation (unlike homologs NtrC1 in *Aquifex* and DctD in *Sinorhizobium*). This is consistent with a low level of constitutive transcription of the *lpxC* gene, with cell cycle dependent transcription induced by phosphorylation. However, while this association is suggestive, additional studies will be needed to verify that *lpxC* is under σ^{54} control through NtrC4, and the range of gene expression under different conditions. Such studies are very challenging with a thermophile (*A.a.* optimum growth is at >90° C) that is difficult to culture.

Further questions

The extent of DNA-bending induced by binding NtrC-like proteins, and its functional role, are not understood. As mentioned earlier, mesophilic NtrC is known to bind cooperatively at two neighboring binding sites, and is believed to induce a significant DNA bend. Whether cooperative binding affects DNA-bending in NtrC or NtrC4 is not known. It is possible that the lack of a large DNA-bend in the NtrC4 complex is a reflection of the thermophilic nature of *A. aeolicus*. A number of indicators suggest that DNA-binding in thermophiles reflects a different thermodynamic landscape than in mesophiles. For example, it has been shown that at high temperatures curved DNA segments with periodic A-tracts straighten.[45] It has also been suggested that a different balance of topoisomerases and DNA-binding proteins is required to keep the DNA in a duplex state above the melting temperature it would have when linear.[46] Finally, it has been observed that thermophiles tend to positively supercoil DNA, as opposed to the more common negative supercoiling observed in eukarya and eubacteria, which may also affect the thermodynamics of DNA binding.[47] There is currently little understanding of how these features might affect the DNA complex formation at the high temperatures that are normal growth conditions for *A. aeolicus* cells (which are >90°C).

NtrC proteins have a very different function in the cell than Fis does, and this is reflected in their DNA binding. Transcriptional activation through σ^{54} requires localization of the protein activator to the correct genes to be transcribed. The occupancy of the activator binding sites in a promoter region largely determines the level of transcription from the corresponding gene when the activating signal for the activator is present (phosphorylation for NtrC4), hence for the NtrC family of proteins selectivity in binding is directly reflected

in selectivity of transcription. It appears that the evolution of Fis from an NtrC-like specific-binding transcription factor into a versatile, low specificity DNA-bending protein occurred with very subtle changes to the number and positioning of specific and nonspecific contacts. Perhaps with new methods being applied to these systems[48] it will become possible to expreimentally address these issues.

MATERIALS AND METHODS

Protein expression, purification

Two constructs of the DNA-binding domain of NtrC4 from *A. aeolicus* (365–442 and 374–442) were cloned, expressed, and purified as previously described.[11]

Crystallography of 4DBD (unbound)

Crystals were grown at room temperature from 30 mg/ml protein in 20% PEG 550 MME, 0.1 M MES pH 6.5 and 0.01 M zinc sulfate. Crystals only formed in the presence of zinc or cadmium. Diffraction data were collected at the Advanced Light Source at Lawrence Berkeley National Laboratory at beamlines 8.2.2 and 8.3.1. A 3 wavelength MAD dataset was collected (Zn peak 1.2829 Å, Zn edge 1.2823 Å, Zn remote 1.2193 Å, with phasing powers of 0.993, 0.918 and 0.859 respectively) from a single zinc-containing crystal (Table 1). The program Elves[49] was used to automatically process the data and solve the structure. Elves uses MOSFLM[50] and SCALA[51, 52, 53] to process and scale the data, and locates heavy atom sites with Shelx[54] and subsequently refines them using MLPHare[55]. Initial phases were calculated with MLPHare. Density modification was done using DM[56] and RESOLVE[57] and the electron density map was calculated with the CCP4 suite[58]. Models were build manually using O[59] and refined in REFMAC5[60] within the CCP4 suite. Refinement was monitored using a 5% free-R sample. Anisotropic B-factors and bulk solvent correction were applied throughout the refinement. TLS refinement was carried out after the free-R set dropped below 30%. Water picking was performed using ArpWarp[61] after the free-R set dropped below 28%. Data were also collected from a cadmium containing crystal, giving an essentially identical structure (data not shown).

NMR Experiments

All NMR experiments were performed at 25°C on uniformly labeled ^{15}N-^{13}C or uniformly ^{15}N-labeled samples at 1 mM in 250 mM NaCl, 50 mM sodium phosphate buffer pH 6.5, 1 mM EDTA, and 10% D_2O. Backbone resonances were assigned by combining HNCA, 3D ^{15}N-NOESY FHSQC and CBCA(CO)NH experiments. Verification of the initial assignment and resolution of ambiguous regions was obtained from HB(CB)HA(CACO)NH, HNCACB and 3D TOCSY-HSQC spectra. The steady state NOE technique was used to obtain ^{1}H-^{15}N NOE values.[62] Two 2D FHSQC type spectra were recorded in the presence and absence of a series of proton presaturation pulses, with mixing time of 100 ms. All data were processed with NMRPipe[63] and analyzed with CARA[64].

Complex preparation for crystallography

The following DNA forward and reverse strands were ordered unpurified from the Keck Facility at Yale University for crystallization trials: 5′-ACTTGCAAATTTGCAAATGCAT

-3′ (forward) and 5′-TGATGCATTTGCAAATTTGCAA -3′ (reverse). DNAs were purified by large gel electrophoresis in 17% 29:1 denaturing polyacrylamide gels in 8 M urea. DNA bands were visualized by DNA shadowing, excised from the gel with a razor, pulverized by extrusion through a syringe tip, and soaked in 25 mL aliquots of TBE 1X buffer with rigorous shaking at 37°C. A total of 150 mL of aliquots were taken over 36 hours. Extracted DNA was desalted with Waters Sep-Pak desalting columns, eluted into 50% methanol, and dried in a speed-vac centrifuge. Strands were resuspended in 50 mM HEPES pH 7.0, 20 mM NaCl, 1 mM EDTA ("DNA-binding buffer"), quantified by UV absorption at 260 nm, combined in 1:1 molar ratios in the presence of 2 mM $MgCl_2$, and annealed in a hot water bath for 3 hours. The resulting DNA was combined in a 1:2 molar ratio with 4DBD monomers in DNA-binding buffer, and run over a Superdex75 size exclusion column equilibrated with the same buffer to reduce excess amounts of free ssDNA, dsDNA, or protein. The eluted complex sample was concentrated with Milipore Centricon concentrators to a concentration of 20–30 mg/ml, and stored at 4°C. Protein concentrations in complex samples were quantified with a Bradford assay.

Crystallography of 4DBD-complex

4DBD was co-crystallized by the hanging-drop method with dsDNA in 30 mM Na-cacodylate pH 6.0, 140 mM spermidine, 50 mM potassium thiocyanate, and 15% (poly)ethylene glycol mw 400, at 20°C, cryoprotected in 20% (w/v) xylitol and frozen in liquid nitrogen until use. X-ray diffraction data were collected at the Advanced Light Source (ALS) beamline 8.3.1 from a single crystal to 3.0 Å resolution at a wavelength of 0.92 Å over a 120° wedge with 1° oscillations. Data were integrated with DENZO and scaled with SCALEPACK as implemented in the HKL2000 suite of software.[65] Initial phases were solved by molecular replacement using the program PHASER.[66] The starting model included two copies of the 4DBD monomer from the free-protein crystal structure, together with a model of ideal 13-mer B-form dsDNA (3D-DART server[67]). Remaining DNA nucleotides were modeled manually using Coot,[68] and refined with PHENIX.[69] The structure was validated using SFCHECK,[70] PROCHECK[71] and MOLPROBITY.[72] Distances between recognition helices were measured by the helix-distance function in Chimera.[73] The structural statistics are shown in Table 1.

Gel-Shift Assays to determine NtrC4 binding site

Upstream activating sequences (UASs) from Aquifex aeolicus contained 250-bp regions upstream from σ^{54}-binding sites, including those upstream of the *dhsU*, *nirB*, *fhp*, *glnB*, *Aq_087*, and *lpxC* genes. UASs were amplified by hot-start PCR, and purified gel band extraction from 2% agarose gels utilizing the Qiagen mini-prep gel purification kit. To determine the promoter regions to which NtrC4 binds, 100 nM of each individual UAS was mixed with varying concentrations of full-length NtrC4 (prepared as previously described[11]) in 25 mM Tris buffer pH 8.0, 150 mM NaCl, 1 mM EDTA, and 10% glycerol, and run on a 1% low-melting-point (LMP) agarose gel at 100V for one hour at room temperature. Bands were visualized by SYBR-gold staining (Invitrogen) with a UV visualization box.

Dnase I footprinting

The upstream activation site (UAS) of the lpxC promoter was inserted into the pMAL-p2X vector (New England Biolabs). The resulting pMAL-p2X derivative was used as template for a DNAseI footprinting assay. 1 nM (final concentration) of supercoiled plasmid was incubated with different concentrations of NtrC4 full-length activator (final concentration 1, 2, 4, 8, 16 or 50 nM). After a 10-min incubation, 2 μg/ml DNase I (from Sigma) was added for 40 seconds. A sample containing only the template plasmid was used as control, for which DNaseI was added for 20 seconds. The reactions were stopped by the addition of 150 ml of PB buffer from Qiagen (chaotropic salts) and purified with the Qiagen PCR Purification Kit. All purified DNA fragments were analysed through a 14-round primer extension using a ^{32}P labeled primer (5′-CGTATAATGTGTGGAATTGTGAGCGG). After the primer extension, all samples were run on 6% PAGE with 1X Tris-buffered EDTA (10 mM Tris buff pH 8.0, 1 mM EDTA) at 32 W for 100 min and radioactive bands were visualized by phosphorimager analysis.

DNA-binding assays to determine equilibrium constants for 4DBD binding to specific sites

DNA forward and reverse strands of 26-mer binding-site variants (mutated residues bolded/underlined: variant_2: CATTTTTTGCAAATTTGCAAATGCAT; variant_3: CATTTTTT**T**CAAATTTGCAAATGCAT; variant_4: CATTTTTTG**T**AAATTTGCAAATGCAT; variant_5: CATTTTTTGCAAATTT**T**CAAATGCAT; variant_6: CATTTTTTGCAAATTTG**T**AAATGCAT; variant_7: CATTTT**C**TGCAAATTTGCA**C**ATGCAT); variant_8: CATTTTTTGCAA**CG**TTGCAAATGCAT were purchased from IDT with HPLC purification and with a fluorescein tag on the 5′-end of the forward strand. Duplexes were annealed by slow cooling in a water bath from 100 C to room temperature in 50 mM phosphate buffer pH 6.5, 5 mM NaCl and 1 mM EDTA, and gel-purified on 20% 70:1 polyacrylamide gels with the Qiagen mini-prep gel extraction kit. Combinations of dimeric 4DBD and different purified duplex DNAs were run on pre-cast Biorad Tris-HCl 4–12% gradient polyacrylamide gels at 80V for 3 hours at 4°C, and bands were visualized and quantified on a Typhoon Trio Imager fluorescence laser scanner (GE Healthcare). Band intensities were plotted versus dimer concentration, and equilibrium binding constants were calculated with a standard two-site Adair model utilizing Prism.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

FUNDING

This work was supported by the National Institutes of Health (grant GM62163); N.K.V. was also supported by a National Science Foundation Graduate Research Fellowship. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. NMR Instrumentation grants from the U.S. Department of Energy, DE FG05-86ER75281, and the National Science Foundation, DMB 86-09305 (500 MHz) and BBS 87-20134 (600 MHz); and NSF (BBS 01-19304) and NIH (RR15756) (800 MHz) are gratefully acknowledged.

The authors appreciate many fruitful discussions with Sydney Kustu, Michaeleen Callahan-Doucleff, and Joseph Batchelor, and the staff of ALS beamline 8.3.1 for help and advice. We dedicate this paper to the memory of Sydney Kustu.

References

1. Ghosh T, Bose D, Zhang X. Mechanisms for activating bacterial RNA polymerase. FEMS Microbiol Rev. 2010; 34:611–627. [PubMed: 20629756]
2. Paget MS, Helmann JD. The sigma70 family of sigma factors. Genome Biol. 2003; 4:203. [PubMed: 12540296]
3. Buck M, Gallegos MT, Studholme DJ, Guo Y, Gralla JD. The bacterial enhancer-dependent σ^{54} (sigma(N)) transcription factor. J Bacteriol. 2000; 182:4129–4136. [PubMed: 10894718]
4. Campbell EA, Westblade LF, Darst SA. Regulation of bacterial RNA polymerase σ factor activity: a structural perspective. Curr Opin Microbiol. 2008; 11:121–127. [PubMed: 18375176]
5. Wigneshweraraj S, Bose D, Burrows PC, Joly N, Schumacher J, Rappas M, Pape T, Zhang X, Stockley P, Severinov K, et al. Modus operandi of the bacterial RNA polymerase containing the σ^{54} promoter-specificity factor. Mol Microbiol. 2008; 68:538–546. [PubMed: 18331472]
6. Barrios H, Valderrama B, Morett E. Compilation and analysis of σ^{54}-dependent promoter sequences. Nucleic Acids Res. 1999; 27:4305–4313. [PubMed: 10536136]
7. da Silva Neto JF, Koide T, Gomes SL, Marques MV. Global gene expression under nitrogen starvation in Xylella fastidiosa: contribution of the σ^{54} regulon. BMC Microbiol. 2010; 10:231. [PubMed: 20799976]
8. Amouyal M. Gene regulation at-a-distance in E. coli: new insights. C R Biol. 2005; 328:1–9. [PubMed: 15714875]
9. Ramos JL, Marques S, Timmis KN. Transcriptional control of the Pseudomonas TOL plasmid catabolic operons is achieved through an interplay of host factors and plasmid-encoded regulators. Annu Rev Microbiol. 1997; 51:341–373. [PubMed: 9343354]
10. Montie TC, Doyle-Huntzinger D, Craven RC, Holder IA. Loss of virulence associated with absence of flagellum in an isogenic mutant of Pseudomonas aeruginosa in the burned-mouse model. Infect Immun. 1982; 38:1296–1298. [PubMed: 6818148]
11. Batchelor JD, Doucleff M, Lee CJ, Matsubara K, De Carlo S, Heideker J, Lamers MH, Pelton JG, Wemmer DE. Structure and regulatory mechanism of Aquifex aeolicus NtrC4: variability and evolution in bacterial transcriptional regulation. J Mol Biol. 2008; 384:1058–1075. [PubMed: 18955063]
12. Pelton JG, Kustu S, Wemmer DE. Solution structure of the DNA-binding domain of NtrC with three alanine substitutions. J Mol Biol. 1999; 292:1095–1110. [PubMed: 10512705]
13. Sallai L, Tucker PA. Crystal structure of the central and C-terminal domain of the σ^{54}-activator ZraR. J Struct Biol. 2005; 151:160–170. [PubMed: 16005641]
14. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM. The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol Rev. 2005; 29:231–262. [PubMed: 15808743]
15. Raetz CR, Whitfield C. Lipopolysaccharide endotoxins. Annu Rev Biochem. 2002; 71:635–700. [PubMed: 12045108]
16. Morett E, Bork P. Evolution of new protein function: recombinational enhancer Fis originated by horizontal gene transfer from the transcriptional regulator NtrC. FEBS Lett. 1998; 433:108–112. [PubMed: 9738943]
17. Dorman CJ. Nucleoid-associated proteins and bacterial physiology. Adv Appl Microbiol. 2009; 67:47–64. [PubMed: 19245936]
18. Ussery D, Larsen TS, Wilkes KT, Friis C, Worning P, Krogh A, Brunak S. Genome organisation and chromatin structure in Escherichia coli. Biochimie. 2001; 83:201–212. [PubMed: 11278070]
19. Hengen PN, Bartram SL, Stewart LE, Schneider TD. Information analysis of Fis binding sites. Nucleic Acids Res. 1997; 25:4994–5002. [PubMed: 9396807]
20. Grove A. Functional evolution of bacterial histone-like HU proteins. Curr Issues Mol Biol. 2011; 13:1–12. [PubMed: 20484776]

21. Cheng YS, Yang WZ, Johnson RC, Yuan HS. Structural analysis of the transcriptional activation region on Fis: crystal structures of six Fis mutants with different activation properties. J Mol Biol. 2000; 302:1139–1151. [PubMed: 11183780]
22. Yuan HS, Wang SS, Yang WZ, Finkel SE, Johnson RC. The structure of Fis mutant Pro61Ala illustrates that the kink within the long alpha-helix is not due to the presence of the proline residue. J Biol Chem. 1994; 269:28947–28954. [PubMed: 7961857]
23. Kostrewa D, Granzin J, Stock D, Choe HW, Labahn J, Saenger W. Crystal structure of the factor for inversion stimulation FIS at 2.0 A resolution. J Mol Biol. 1992; 226:209–226. [PubMed: 1619650]
24. Yuan HS, Finkel SE, Feng JA, Kaczor-Grzeskowiak M, Johnson RC, Dickerson RE. The molecular structure of wild-type and a mutant Fis protein: relationship between mutational changes and recombinational enhancer function or DNA binding. Proc Natl Acad Sci U S A. 1991; 88:9558–9562. [PubMed: 1946369]
25. Pan CQ, Finkel SE, Cramton SE, Feng JA, Sigman DS, Johnson RC. Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. J Mol Biol. 1996; 264:675–695. [PubMed: 8980678]
26. Thompson JF, Landy A. Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes. Nucleic Acids Res. 1988; 16:9687–9705. [PubMed: 2972993]
27. Revet B, Brahms S, Brahms G. Binding of the transcription activator NRI (NTRC) to a supercoiled DNA segment imitates association with the natural enhancer: an electron microscopic investigation. Proc Natl Acad Sci U S A. 1995; 92:7535–7539. [PubMed: 7638226]
28. Ray P, Smith KJ, Parslow RA, Dixon R, Hyde EI. Secondary structure and DNA binding by the C-terminal domain of the transcriptional activator NifA from Klebsiella pneumoniae. Nucleic Acids Res. 2002; 30:3972–3980. [PubMed: 12235381]
29. Stella S, Cascio D, Johnson RC. The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. Genes Dev. 2010; 24:814–826. [PubMed: 20395367]
30. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr Sect D: Biol Crystallogr. 2004; 60:2256–2268. [PubMed: 15572779]
31. Hertz GZ, Stormo GD. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics. 1999; 15:563–577. [PubMed: 10487864]
32. Studholme DJ, Wigneshwereraraj SR, Gallegos MT, Buck M. Functionality of purified σ^N (σ^{54}) and a NifA-like protein from the hyperthermophile Aquifex aeolicus. J Bacteriol. 2000; 182:1616–1623. [PubMed: 10692367]
33. Haran TE, Mohanty U. The unique structure of A-tracts and intrinsic DNA bending. Q Rev Biophys. 2009; 42:41–81. [PubMed: 19508739]
34. Beveridge DL, Dixit SB, Barreiro G, Thayer KM. Molecular dynamics simulations of DNA curvature and flexibility: helix phasing and premelting. Biopolymers. 2004; 73:380–403. [PubMed: 14755574]
35. Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc Natl Acad Sci U S A. 1998; 95:11163–11168. [PubMed: 9736707]
36. Lavery R, Moakher M, Maddocks JH, Petkeviciute D, Zakrzewska K. Conformational analysis of nucleic acids revisited: Curves+ Nucleic Acids Res. 2009; 37:5917–5929. [PubMed: 19625494]
37. Skoko D, Yoo D, Bai H, Schnurr B, Yan J, McLeod SM, Marko JF, Johnson RC. Mechanism of chromosome compaction and looping by the Escherichia coli nucleoid protein Fis. J Mol Biol. 2006; 364:777–798. [PubMed: 17045294]
38. North AK, Kustu S. Mutant Forms of the Enhancer-Binding Protein NtrC can Activate Transcription from Solution. J Mol Biol. 1997; 267:17–36. [PubMed: 9096204]
39. Vidangos N, Maris AE, Young A, Hong E, Pelton JG, Batchelor JD, Wemmer DE. Structure, function, and terthering of DNA-binding domains in s(54) transcriptional activators. Biopolymers. 2013; 99:1082–1096. [PubMed: 23818155]

40. Ninfa AJ, Reitzer LJ, Magasanik B. Initiation of transcription at the bacterial glnAp2 promoter by purified E. coli components is facilitated by enhancers. Cell. 1987; 50:1039–1046. [PubMed: 3304660]
41. De Carlo S, Chen B, Hoover TR, Kondrashina E, Nogales E, Nixon BT. The structural basis for regulated assembly and function of the transcriptional activator NtrC. Genes Dev. 2006; 20:1485–1495. [PubMed: 16751184]
42. Santero E, Hoover TR, North AK, Berger DK, Porter SC, Kustu S. Role of integration host factor in stimulating transcription from the sigma 54-dependent nifH promoter. J Mol Biol. 1992; 227:602–620. [PubMed: 1404379]
43. Barb AW, Zhou P. Mechanism and inhibition of LpxC: an essential zinc-dependent deacetylase of bacterial lipid A synthesis. Curr Pharm Biotechnol. 2008; 9:9–15. [PubMed: 18289052]
44. Francke C, Groot Komelink T, Hagemeijer Y, Overmars L, Sluijter V, Moezelaar R, Siezen RJ. Comparative analyses imply that the enigmatic Sigma factor 54 is a central controller of the bacterial exterior. BMC Genomics. 2011; 12:385. [PubMed: 21806785]
45. Diekmann S. Temperature and salt dependence of the gel migration anomaly of curved DNA fragments. Nucleic Acids Res. 1987; 15:247–265. [PubMed: 3029673]
46. Guagliardi A, Napoli A, Rossi M, Ciaramella M. Annealing of complementary DNA strands above the melting point of the duplex promoted by an archaeal protein. J Mol Biol. 1997; 267:841–848. [PubMed: 9135116]
47. Napoli A, Zivanovic Y, Bocs C, Buhler C, Rossi M, Forterre P, Ciaramella M. DNA bending, compaction and negative supercoiling by the architectural protein Sso7d of Sulfolobus solfataricus. Nucleic Acids Res. 2002; 30:2656–2662. [PubMed: 12060682]
48. Friedman LJ, Gelles J. Mechanism of transcriptional initiation at an activator-dependent promoter defined by single-molecule observation. Cell. 2012; 148:679–689. [PubMed: 22341441]
49. Holton J, Alber T. Automated protein crystal structure determination using ELVES. Proc Natl Acad Sci U S A. 2004; 101:1537–1542. [PubMed: 14752198]
50. Leslie, AGW. Crystallographic Computing: From Chemistry to Biology. Moras, D.; Podjarny, AD.; Thiery, JC., editors. Oxford University Press; Oxford: 1991. p. 27-38.
51. Kabsch W. Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. J Appl Crystallogr. 1988; 21:916–924.
52. Evans, PR. Data reduction. In: Sawyer, L.; Issacs, NW.; Bailey, S., editors. Proceedings of the CCP4 Study Weekend: Data Collection and Processing. Warrington: Daresbury Laboratory; 1993. p. 114-122.
53. Evans, PR. Proceedings of the CCP4 Study Weekend: On Recent Advances in Phasing. Warrington: Daresbury Laboratory; 1997. Scaling of MAD Data.
54. Sheldrick GM. Phase Annealing in Shelx-90 - Direct Methods for Larger Structures. Acta Crystallogr Sect A. 1990; 46:467–473.
55. Otwinowski, Z. Maximum likelihood refinement of heavy atom parameters. In: Wolf, W.; Evans, PR.; Leslie, AGW., editors. Proceedings of the CCP4 Study Weekend Isomorphous Replacement and Anomalous Scattering. Warrington: Daresbury Laboratory; 1991. p. 80-86.
56. Cowtan K. “dm”: An automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography. 1994; 31:34–38.
57. Terwilliger TC. Reciprocal-space solvent flattening. Acta Crystallogr Sect D: Biol Crystallogr. 1999; 55:1863–1871. [PubMed: 10531484]
58. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr Sect D: Biol Crystallogr. 2011; 67:235–242. [PubMed: 21460441]
59. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved Methods for Building Protein Models in Electron-Density Maps and the Location of Errors in These Models. Acta Crystallogr Sect A. 1991; 47:110–119. [PubMed: 2025413]
60. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr Sect D: Biol Crystallogr. 1997; 53:240–255. [PubMed: 15299926]

61. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. Nat Struct Biol. 1999; 6:458–463. [PubMed: 10331874]
62. Farrow NA, Zhang O, Forman-Kay JD, Kay LE. A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chemical exchange rates of systems in slow equilibrium. J Biomol NMR. 1994; 4:727–734. [PubMed: 7919956]
63. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995; 6:277–293. [PubMed: 8520220]
64. Keller, RLJ. Thèse de doctorat, ETH Zurich Thesis No 15947. Switzerland: 2004. Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment.
65. Otwinowski, ZMW. Processing of X-ray diffraction data collected in oscillation mode. In: Carter, JRMS., editor. Methods in enzymology. Vol. 276. Academic Press; New York: 1997. p. 307-326.
66. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007; 40:658–674. [PubMed: 19461840]
67. van Dijk M, Bonvin AM. 3D-DART: a DNA structure modelling server. Nucleic Acids Res. 2009; 37:W235–239. [PubMed: 19417072]
68. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr Sect D: Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]
69. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr Sect D: Biol Crystallogr. 2010; 66:213–221. [PubMed: 20124702]
70. Vaguine AA, Richelle J, Wodak SJ. SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta Crystallogr Sect D: Biol Crystallogr. 1999; 55:191–205. [PubMed: 10089410]
71. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR. 1996; 8:477–486. [PubMed: 9008363]
72. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr Sect D: Biol Crystallogr. 2010; 66:12–21. [PubMed: 20057044]
73. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004; 25:1605–1612. [PubMed: 15264254]

HIGHLIGHTS

- Transcription initiation by bacterial σ^{54}-polymerase requires an upstream activator
- The structure of an activator's DNA-binding domain, with and without DNA, was solved
- The activator's DNA-binding domain is homologous to a multipurpose protein, Fis
- Fis and NtrC4 contact DNA similarly, but Fis is more reliant on nonspecific contacts
- These subtle differences explain the proteins' divergent functions.

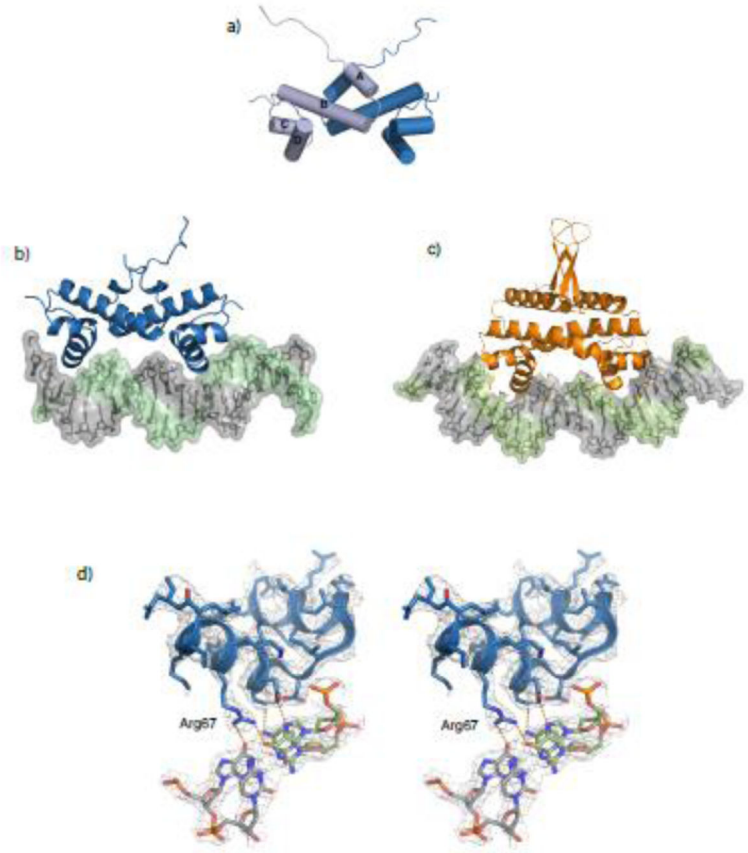

Figure 1.

Structure of free and bound forms of 4DBD. a) Cylinder view of free 4DBD crystal structure (3E7L) showing the two symmetrical dimers in light and dark blue; helix labeling convention A–D is shown. N-terminal segments not seen in the density were added in a random, extended conformation to indicate the construct present in the crystal. b) Cartoon view of 4DBD bound to variant_1 DNA (4FTH). Forward strand is gray, reverse strand is green. c) structure of Fis bound to DNA,[29] d) stereo image cutaway of structure and simulated annealing composite omit (SA-omit) 3Fo-2Fc map to highlight specific contacts.

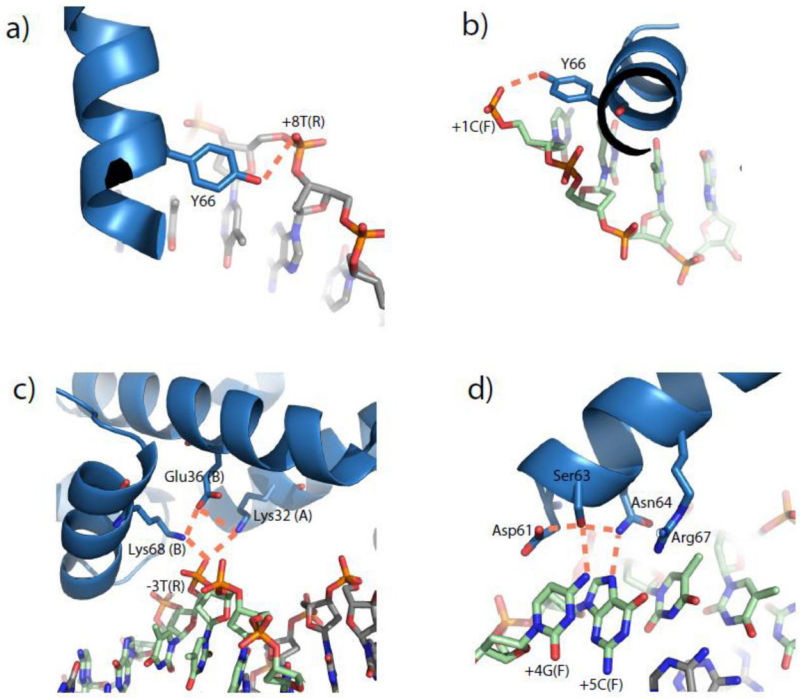

Figure 2.

Cutaway views of nonspecific interactions. To use the same numbering convention in the next two figures, the DNA is labeled by relative position (see Figure 3) and strand. **a)** Y66 interaction with internal backbone phosphate, contrasted with **b)** the other monomer's Y66 interaction with the first resolvable base pair backbone near the edge of the DNA-binding site, **c)** hydrogen bonding ring at +3T, **d)** Asp61, Ser63, Asn64 hydrogen bond network. Forward and Reverse DNA strands are indicated by (F) and (R), and in **c)**, residues on different 4DBD monomers are distinguished by labeling them (A) and (B).

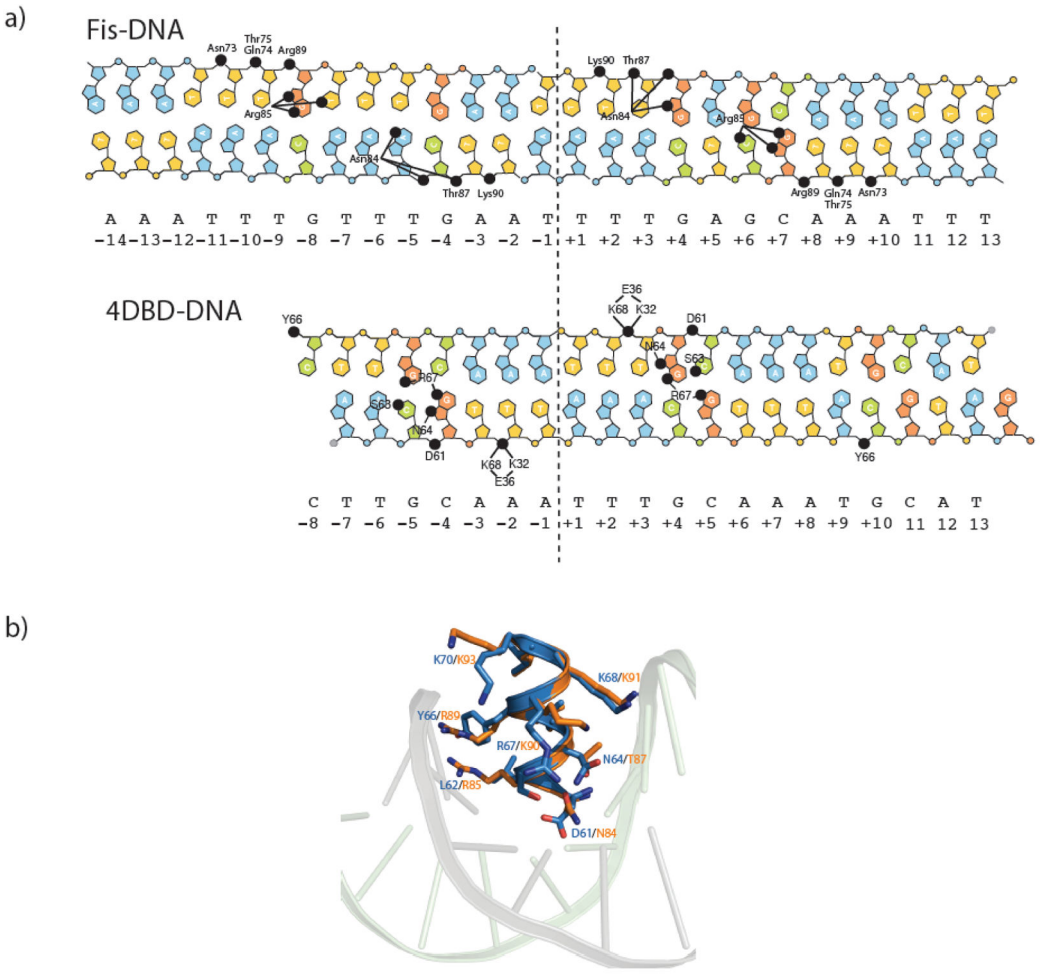

Figure 3.

a) DNA-binding schematics showing specific and nonspecific contacts of Fis and 4DBD with base pair numbering convention. Note that 4DBD residue −9A on the forward strand is omitted from this figure, as it was not visible in the density. b) cutaway view of the D helices of 4DBD (blue) and Fis (orange) to highlight the overlapping residues (NtrC4: K68, Y66, D61, N64).

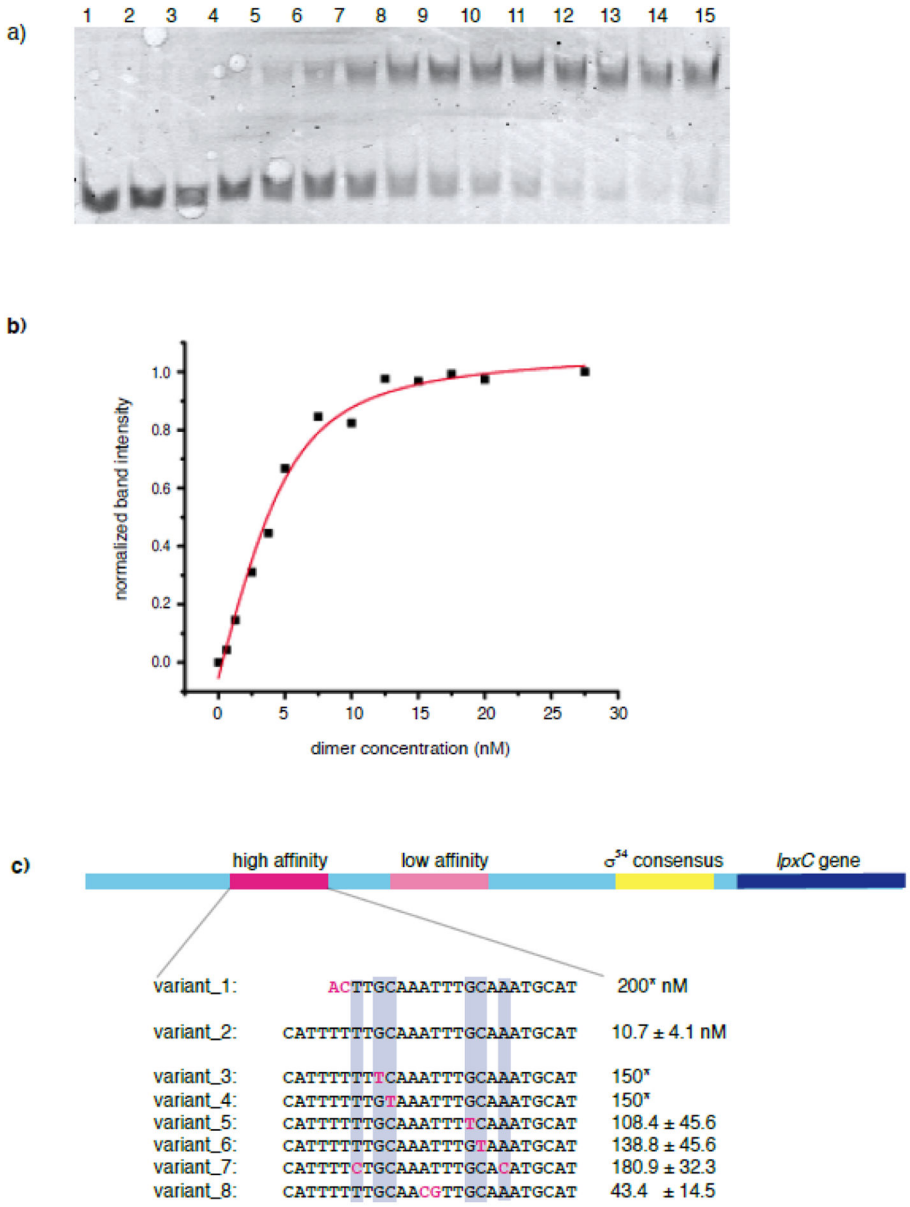

Figure 4.

a) A gel shift binding-affinity assay to determine the K_d of 4DBD in complex with variant_2 (high affinity, native extended site). Each lane contains 5 nM fluorscein-labeled DNA and the following dimeric protein concentrations: 0, 0.6, 1.3, 2.5, 3.8, 5.0, 7.5, 10.0, 12.5, 15.0, 17.5, 20.0, 22.5, 25.0, 27.5 nM. **b)** Plot of intensity vs. protein concentration from data in a). Red line is fit of data to the standard Adair two-site binding curve, as described in the Methods section. **c)** Schematic of promoter and UAS and summary of binding site mutations tested and their resulting K_d values. Values with an asterisk showed non-two-state behavior, the value given corresponds to 50% of the DNA shifted.

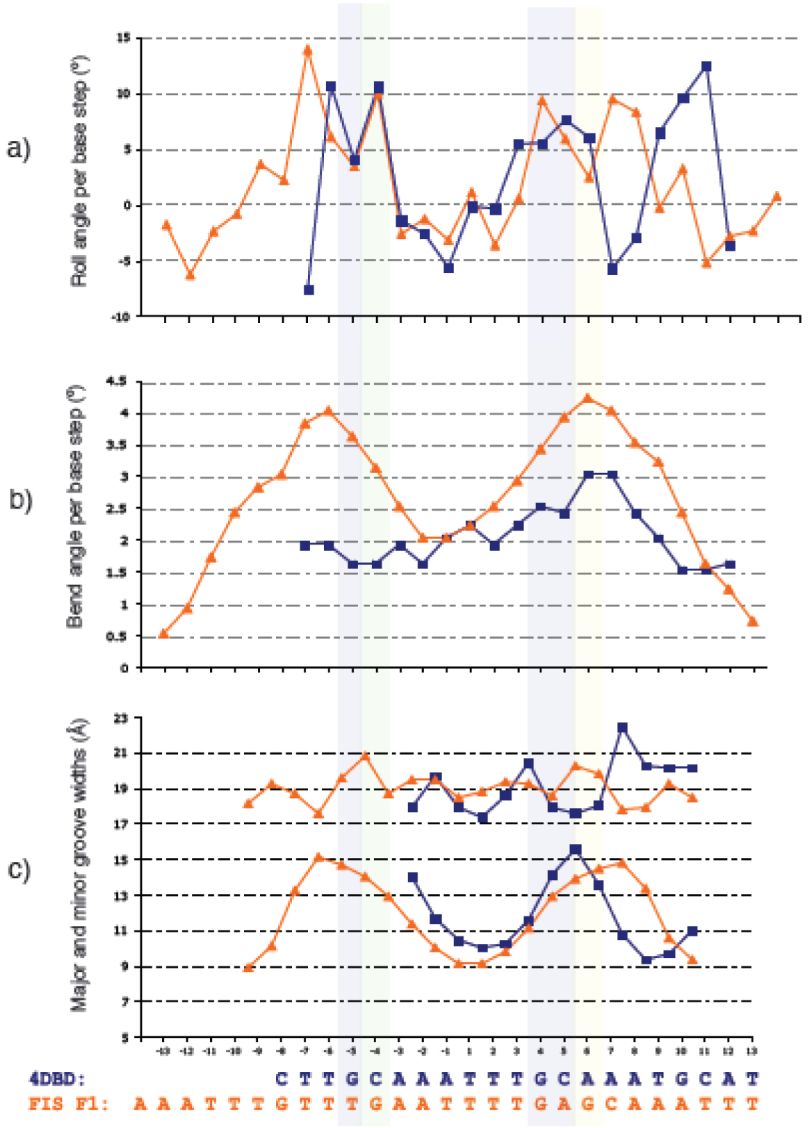

Figure 5.

Comparison of DNA distortion in Fis (orange) and 4DBD(blue). **a)** Roll angles, **b)** bend angles, **c)** major groove (upper lines) and minor groove (lower lines) widths per base step.

Table 1

X-ray crystal structure statistics.

	4DBD	4DBD complex
Data Collection		
Space group:	$P2_1$	$P2_1$
Unit Cell parameters		
a, b, c	41.24, 55.87, 62.34	67.4, 28.8, 74.9
α, β, γ	90.00, 93.39, 90.00	90, 94.6, 90
Resolution (Å)	41.17-2.25 (2.43-2.25)	48.1-3.0 (3.1-3.0)
Unique Reflections	13,286(1273)	5,902 (454)
Completeness (%)	98.1% (94.2)	96.8% (76.2)
Multiplicity	3.4 (3.0)	3.3 (2.3)
I/σ_1	16.5(3.9)	15.5(2.5)
R_{merge}	0.062 (0.062)	0.063(0.31)
Refinement		
R_{work}/R_{free} (%)	20.3/24.3	24.3 / 26.4
Number of atoms		
Protein	2046	1106
DNA	N/A	861
B-factors		
Protein	75.2	95.4
DNA	N/A	112.6
RMSDs		
bond lengths(Å)	0.004	0.004
bond angles (°)	0.701	0.802

Values in parentheses correspond to the highest resolution shell.