Published in final edited form as:

Biopolymers. 2013 December ; 99(12): 1082–1096. doi:10.1002/bip.22333.

Structure, function, and tethering of DNA-binding domains in σ^{54} transcriptional activators

Natasha Vidangos¹, Ann E. Maris¹, Anisa Young¹, Eunmi Hong¹, Jeffrey G. Pelton¹, Joseph D. Batchelor¹, and David E. Wemmer¹

¹Department of Chemistry and QB3 Institute, University of California, Berkeley CA 94720-1460

Abstract

We compare the structure, activity and linkage of DNA binding domains from σ^{54} transcriptional activators, and discuss how the properties of the DNA binding domains and the linker to the neighboring domain are affected by the overall properties and requirements of the full proteins. These transcriptional activators bind upstream of specific promoters that utilize σ^{54} -polymerase. Upon receiving a signal the activators assemble into hexamers, which then, through ATP hydrolysis, drive a conformational change in polymerase that enables transcription initiation. We present structures of the DNA-binding domains of activators NtrC1 and Nlh2 from the thermophile *A. aeolicus*. The structures of these domains, and their relationship to other parts of the activators are discussed. These structures are compared with previously determined structures of the DNA-binding domains of NtrC4, NtrC, ZraR, and FIS. The N-terminal linkers that connect the DNA-binding domains to the central domains in NtrC1 and Nlh2 were studied and found to be unstructured. Additionally, a crystal structure of full-length NtrC1 was solved, but density for the DNA-binding domains was extremely weak, further indicating that the linker between ATPase and DNA binding domains functions as a flexible tether. Flexible linking of ATPase and DNA binding domains is likely necessary to allow assembly of the active hexameric ATPase ring. The comparison of this set of activators also shows clearly that strong dimerization of the DNA binding domain only occurs when other domains do not dimerize strongly.

INTRODUCTION

Regulation of gene transcription is a very important mechanism mediating bacterial cell responses to environmental cues. In bacteria RNA polymerase (RNAP), the key enzyme that transcribes DNA into mRNA, is a five-subunit core enzyme that associates with a variable sixth subunit, the σ -factor, to form the holoenzyme (1,2). σ -factors convey promoter specificity to RNAP through sequence specific DNA binding, and are involved in a number of mechanisms for regulating the transcriptional activity. Control through the σ -factor enables the transcription of different gene classes to be regulated by different regulatory pathways (3,4). σ -factors can be clustered into two structurally unrelated families: σ^{70} and σ^{54} . σ -factors of the σ^{70} class are abundant, with many variants usually present in cells, responding to changing conditions (5). In contrast, most but not all bacteria (6) produce a σ^{54} protein, and there is usually just one σ^{54} gene per genome. σ^{54} commonly regulates transcription of genes associated with processes that require a rapid response with a wide dynamic range, such as starvation responses, flagellar motility, protection from alkaline and high-pressure environments, and expression of virulence factors (5,7,8).

Although all σ factors play an analogous role in targeting promoters, and regulating transcription, the mechanisms of the σ^{70} and σ^{54} classes are quite different. For σ^{70} once the holopolymerase binds at a promoter opening of the DNA initiation of transcription occurs rapidly. Much of the regulation occurs by inhibiting formation of the polymerase-promoter complex, for example by tying up the σ -factor in complexes with anti- σ proteins, or

blocking the promoter with a repressor protein. In σ^{54} -dependent transcription there is a fundamental difference in that σ^{54} factor alone cannot open the promoter DNA, and hence cannot initiate transcription without the action of another protein. Neither assembly of holo-polymerase with σ^{54} , nor binding at the promoter is regulated. For σ^{54} -polymerase to open the promoter, the action of a transcriptional activator is required. The activator proteins bind upstream of the transcription initiation site, in a region called the upstream activating sequence (UAS) or enhancer element. They undergo a regulated assembly to form an oligomer (9–11), which then contacts σ^{54} and uses ATP hydrolysis to drive a conformational change that enables promoter opening, Figure 1.

The usual architecture of σ^{54} transcriptional activators includes domains with three distinct functions: N-terminal regulatory (R), central oligomerization/AAA+ATPase (C), and C-terminal helix-turn-helix DNA-binding domain (D or DBD) (12). The regulatory function may be associated with one or more structural domains. The activators are dimers in the unactivated state, and normally two dimers are bound at specific target DNA sequences in the UAS. Different activators have distinct DNA binding specificities so that they act only on a specific subset of σ^{54} -polymerase transcribed genes. Upon receiving an activating signal a conformational change occurs in the regulatory domain(s) (signals are commonly phosphorylation of a receiver domain or ligand binding by a GAF domain), the dimers oligomerize into hexamers. The hexamers contact the σ^{54} -RNAP holoenzyme (13–15), hydrolyze ATP, and thereby drive a conformational change in the polymerase-promoter complex that enables opening of the promoter and initiation of transcription (16). A number of activators (also known as enhancer binding proteins, or EBPs) have been studied structurally, including NtrC (11, 17), DctD (18), PspF (19), and ZraR (20). Structures have been determined for several individual domains, and some domain combinations, and have provided insight into the processes of regulation and oligomerization (21). A typical transcription initiation process is shown schematically in Figure 1. In this work we focus on the DNA binding domains, considering those from several activators and compare them in light of structures and behavior of other linked domains.

The first structure of a DNA binding domain from an activator was that from NtrC - Nitrogen Regulatory Protein C (22), which was also the first activator discovered (23, 24). NtrC is required for transcription of many genes involved in nitrogen starvation response in enteric bacteria, and is regulated by phosphorylation of its receiver domain by NtrB, a histidine kinase. The structure of the linked central and DNA binding domains from ZraR was reported a few years later (20). ZraR has regulatory receiver domains, like NtrC, however little is known about its biochemistry. The observed assembly of the N-terminal truncation (missing just the receiver domain) into active hexameric ATPase rings shows that its receiver domains repress activity. Domains from several activators from *Aquifex aeolicus* have also been characterized. *Aquifex* is an extremely thermophilic bacterial species (optimum growth at over 90° C), with one σ^{54} gene and six transcriptional activator genes (25). Three of the activators have receiver domains regulating them, and hence are orthologs of NtrC, designated NtrC1, NtrC3 and NtrC4. Two others have GAF domains, and are thus orthologs of NifA (26), designated Nlh1 and Nlh2 (Nif-like homolog) and a third has a PAS domain. The structure of the DNA binding domain alone from NtrC4 was reported (27), and recently also in complex with a target DNA oligomer (28). Here we describe structures of the DNA binding domains of NtrC1 and Nlh2, and a structure of full length NtrC1. These structures, together with those previously reported for other activator domains, support the view that while all of the DNA binding domains bind DNA as dimers the contribution of the DNA binding domain to overall protein dimerization is highly variable and coupled to the behavior of other parts of the protein. Further, the DNA binding domains are flexibly linked to the ATPase domains, a property likely needed to accommodate reorganization of the ATPase domains into oligomers upon activation without disrupting DNA binding.

RESULTS

The sequences of the linker-DBD regions from the σ^{54} activators that are discussed in this work are shown aligned in Figure 2. As is the case for almost all activators, the DNA binding domain is the C-terminal segment of the protein and follows the oligomerization/ATPase domain, connected by a linker that is highly variable in both sequence and length. Constructs were made that contained a truncated and a full linker with the DNA binding domain for NtrC1, and the full linker of Nlh2 with its DNA binding domain. Previously studied NtrC and NtrC4 DBDs also had the linkers included in the constructs used. The structure of ZraR reported by Sallai & Tucker (20) was for a two domain construct containing the full ATPase and DNA binding domains but missing the N-terminal receiver domain.

Structure and DNA binding of NtrC1-DBD

The NtrC1 DBD was made as long (including the linker region, residues 373–439) and short (deleting the linker region, residues 392–439) versions. Both constructs were studied using heteronuclear NMR, which showed that they contained a common folded segment, and that the linker region had chemical shifts corresponding to random coil structure. The shorter of the NtrC1 constructs, NtrC1-sDBD, lacking the linker to the ATPase domain, was successfully crystallized. Diffraction data extending to 1.3 Å were collected at the Advanced Light Source at LBNL. Phasing was done using molecular replacement using the NtrC4 DBD, and the structure was refined. The NtrC1-sDBD crystallized with 8 molecules in the asymmetric unit, with symmetry related pairs organized as dimers that are quite similar to those of NtrC, NtrC4 and ZraR. Structural NMR data (NOEs, C_α chemical shifts, and H_α - H_N coupling constants) are completely consistent with the three helix fold found in the crystal structure, and will not be discussed further. Ultracentrifugation and light scattering indicate that the NtrC1-DBD constructs studied are monomers in solution in spite of its dimerization observed in crystals. Gel shift measurements of DNA binding were done using DNA oligomers containing sequences from the high affinity binding site identified in previous footprinting, CTTTGGCCAAAATGGCAAAAG, with typical inverted repeat elements (29). Both constructs lead to a shift, however when an undecamer DNA with a single half site was used, CTTTGGCCAAAG, binding was not detected for the NtrC1-sDBD but the longer version did bind.

Structure of Nlh2-DBD

The structure of the Nlh2-DBD construct (including residues 365–442 encompassing the full linker to the ATPase domain), was solved using multinuclear 3D NMR. Light scattering and ultracentrifugation experiments verified that this construct was predominantly monomeric in solution. Assignments used standard backbone-based magnetization transfer in H-C-N triple resonance experiments, which were then extended to the sidechains using ^{13}C -TOCSY and NOESY data. Distance constraints were generated from NOEs, and then were used to calculate a family of structures using CYANA. The final set structures superimpose with an 0.72 Å RMSD spread. Lower linewidths for resonances from the linker region in ^1H - ^{15}N HSQC spectra indicated that this part of the protein has mobility independent of the folded DBD. The chemical shifts are consistent with random coil structure and no long range NOEs were observed from linker amino acids.

Structure of full-length NtrC1

To determine the relationship between the regulatory/ATPase domains of NtrC1, which were described previously (10), and the DNA binding domains crystallization trials were done with full-length NtrC1. Crystals were obtained that diffracted to 2.2 Å resolution, and the structure was solved using molecular replacement with the previously determined

regulatory/ATPase structure (PDB: 1NY5). Examination of the full length NtrC1 structure showed that a L183S mutation had occurred, and that a disulfide between loop residues Cys158 and Cys161 had formed. These cause minor structural changes in the receiver and ATPase domains relative to the earlier structure, however these local differences were small and will not be discussed further. Only very weak and diffuse density was observable in the region where the DNA binding domain must be (deduced by constraints of linker length and available volume in the crystal). It was not possible to interpret any of the density with a structural model.

Dimerization interfaces: NtrC1 and Nlh2 DBDs

The NtrC1-DBD dimerization interface seen in the crystal structure is quite small, the main contribution coming from an Ile396-Ile396 interaction on the crossed B helices, and a limited number of van der Waals interactions at the base of the three-helix bundle junction that bury only $\sim 475 \text{ \AA}^2$ per monomer. Given the small buried hydrophobic surface area it is not surprising that dimerization is weak, and that predominantly monomer is seen in solution. To investigate potential interactions if a dimer interface formed for Nlh2-DBD, the structure of the symmetric dimer was modeled by docking two monomers together using the program HADDOCK (30, 31). The lowest energy model of the Nlh2-DBD dimer (with a HADDOCK score of -97 ± 7) overlays closely with the NtrC1-DBD structure, giving an r.m.s.d. of $2.92 \text{ \AA}$. The only hydrophobic residues on the B helix that are in the dimerization interface region are F449, I450, and T453. In this model, the I450 residues on opposing B helices, as in NtrC1-DBD, form a van der Waals contact, the two T453 residues form a hydrogen bond to the carbonyl in the backbone of F449, and additional inter-monomer hydrophobic contacts are formed between the F449 side chains with the LGY sequence at the three-helix bundle junction, burying a total of $\sim 500 \text{ \AA}^2$ per monomer, again consistent with a weak dimerization. The structure reported for ZraR (20) has three dimers of DNA binding domains that are very similar to that predicted for NtrC1-DBD and modeled for Nlh2. These data support the view that DNA recognition in NtrC1, Nlh2 and ZraR occurs with dimeric structures. The high affinity DNA binding site for NtrC1 that was identified is an inverted repeat, a likely feature for dimeric Nlh2 and ZraR target sequences as well.

DISCUSSION

The function of the DBDs in transcriptional activation has been studied biochemically, showing that an important function of the NtrC DBD is to localize NtrC oligomers near the σ^{54} promoter, increasing the probability of an interaction with RNAP-holoenzyme (32). A mutant form of NtrC with three alanine substitutions that abolish DNA-binding (NtrC-3Ala) was found to have 5000-fold reduced ability to activate transcription *in vitro*, but that activity could be recovered to wild-type levels by increasing the activator concentration 50-fold (9). While localization is important, close proximity of the DNA-binding site to the promoter with σ^{54} -RNAP is not vital since NtrC can still activate transcription when the NtrC binding sites were moved one kilobase upstream of their native binding site (33), or when attached to separate but concatenated strands of circular DNA (15). These observations argue that DNA binding by the activator is important primarily to generate a high local concentration of the activated ATPase so that it can find and act on the appropriate σ^{54} -RNAP-promoter complexes (i.e. those at the regulated genes) to remodel them and allow promoter opening. The combination of new structures of activator DNA binding domains, and those already published, provide insights into how these function in creating the appropriate localization with sufficient flexibility to allow interconversion between inactive and activated assemblies.

The DNA recognition element is the same in all studied members of this family

With structures of five DNA binding domains from σ^{54} transcriptional activators available it is possible to compare structural features. In the comparisons we also include FIS (Factor for Inversion Stimulation), which evolved from NtrC by losing the regulatory and ATPase domains (34), although its DNA binding is less sequence specific than that of the characterized activators (and probably all activators). In Figure 3 the structure of the first activator DNA binding domain determined, that from NtrC, is shown in cylinder format for reference. Helices (from the N terminus to the C terminus) are labeled A (dimer buttress), B (dimerization helix), C (first helix of the helix-turn-helix), and D (recognition helix), in keeping with general naming conventions for helix-turn-helix proteins (35–37). As will be discussed further below, the dimer buttress helix is not always present. Figure 3 also shows a superposition of monomer structures, generated using just helices B–C–D which are always present, to superimpose them. This helix-turn-helix containing element is very similar in all five activators, giving a pairwise average r.m.s.d. for the structures of 1.4 Å. This is a very common fold element in bacterial DNA binding proteins (35), most of which are dimeric and recognize approximately palindromic sequences. Within this part of the structure there are only slight differences, for example some slight bends/kinks in helices. The length of the B helix also varies: NtrC1-DBD (16 residues); Nlh2-DBD (18 residues); NtrC4-DBD (20 residues); and NtrC-DBD (21 residues). Helix B is central to dimerization, although as discussed further below probably the presence of the A helix is even more important for dimer stability.

The structures of the NtrC1 and Nlh2 as dimers are shown in Figure 4, together with previously published structures of DNA binding domains of NtrC, NtrC4 and ZraR, and also FIS. For Nlh2 the dimer structure shown is the model generated by docking monomers, as discussed in Results, since this domain alone does not dimerize substantially in solution. NtrC4- and NtrC-DBDs, and FIS, have the additional N-terminal A-helix that contributes to the dimerization interface that is not present in the monomeric Nlh2 structure, and all form stable dimers in solution.

NMR data show that the three-helix fold that contacts DNA is present in the NtrC1-DBD in solution, but the B-helix does not seem to extend as far toward the N-terminus as it does in the crystal structure of the dimer. Further, the NMR data do not support the presence of a stable A-helix that is a well-folded part of the structure in NtrC and NtrC4. These observations suggest that the contacts between the B helices seen in the crystal structure are the primary region of interaction upon DNA binding, as discussed further below.

The ZraR-DBD is most similar to NtrC1-DBD. Only half of the ZraR-DBDs present are actually visible in the density in the reported crystal structure (the remainder are presumably present but disordered), and no density is visible connecting the DBD to the ATPase domains, which are assembled as active hexamers in the structure. This structure further supports the view that although the dimer interfaces in NtrC1, Nlh2 and ZraR are quite small, that they can be sufficient to induce formation of a well defined dimer when at high concentration either through crystallization, dimerization of neighboring domains or DNA binding. The relative positions of the recognition helices are fairly similar in NtrC1 and ZraR (though their orientation is somewhat different), and also to the model for Nlh2. Because the dimerization interface is relatively small for this set of proteins it is likely that the positions of their recognition helices might be more pliable than in those proteins with a buttressing A-helix (NtrC, NtrC4 and FIS).

Variations in dimer structures

More substantial differences between structures are visible when the full dimer structures are considered, Figure 4. In all cases studied the activator DNA binding domains interact with DNA as dimers at sequences with inverted repeat elements, which are very common in elements recognized by bacterial DNA binding proteins. It has been found that the enhancer DNA sequences bound by the activators generally come as pairs of dimer binding sites, in fairly close proximity to one another, facilitating oligomerization upon activation of the regulatory domain (9, 29, 38). Most of the variability in structures of dimeric activator DBDs arises from differences in the contacts of the B helices, both in position and in orientation, producing somewhat different relative positions of the recognition helices. When the full dimers are superimposed the average RMSD between structures increases to an average 2.70 Å. In NtrC, NtrC1, and NtrC4, the angles formed by the B helices are 14°, 33°, and 42°, and minimum distances between the central axis of the recognition helices are 18 Å, 22 Å, and 28 Å, respectively. This suggests that variation in the hydrophobic packing of the dimerization interface results in alterations of the length of DNA contacted by the DNA-binding domain, and probably concomitantly in the spacing of the bases to which specific contacts are made. Differences in dimer contacts also cause variation in the angles formed by the D helices in contacting DNA (Figure 4, side view). At present there are only DNA bound structures for NtrC4-DBD, with fairly high DNA binding specificity, and FIS, with low specificity, which were compared in detail (28). Having just two complexes leaves unclear whether the structural differences in the DBDs alone may lead to differences in how contacts to DNA are made. Features of the DBD structures and amino acid sequences are summarized in Table 1.

The binding sites for NtrC1 were mapped with footprinting using the full length proteins (29). Gel shift studies using the NtrC1 DBD alone verified binding of both the short and long versions to the footprinted sequence, CTTTGGCCAAAATGGCAAAAG. Binding to a half site, CTTTGGCCAAAG occurs with low affinity for the long construct, but is not measurable for the shorter version. The fact that the shorter construct does bind the DNA with the full site indicates that there must be interaction between the two monomers binding, and supports the view that there is an active dimerization interface within the DBD. It is probably contact between B helices observed in the crystal structure of the DBD, which helps position the recognition helices enhancing DNA binding affinity as well as contributing a direct inter-protein interaction. The very similar structure of the ZraR DBD dimer suggests that it would have analogous behavior. Unfortunately the target DNA sequences for ZraR and Nlh2 have not been identified precluding specific binding studies.

Primary dimerization determinants of full length σ^{54} activators vary

Dimerization of activators plays two important roles in function. First, dimeric DBDs can efficiently achieve sequence specific binding with high affinity. Second, activator dimerization serves to achieve assembly of the hexameric ATPase rings with reduced molecularity (three dimers needed rather than six monomers), facilitating assembly at *in vivo* concentrations.

For NtrC the DBD is the dimerization determinant for the full-length protein, and deletion of this domain produces monomeric protein with reduced ATPase activity (39). In their unactivated state the NtrC receiver domains are monomeric, and do not interact with other parts of the protein. The ATPase domains do not interact strongly enough to assemble into hexamers, and hence contribute relatively little to dimerization of the inactive protein. These observations are consistent with the positive regulation mechanism, wherein activation of the receiver domain by phosphorylation causes it to bind to a neighboring ATPase domain to structurally stabilize the hexameric ATPase ring (11). The weak interactions of regulatory

and ATPase domains in the unactivated states leaves just the DNA binding domain for dimerization.

For NtrC1 the unactivated regulatory-ATPase domain combination, without the DBDs, are inherently dimeric (10). The unactivated R domains alone are also dimeric, with a crossed helical pair extending from a core dimer interface in the receiver domain. This pair of helices docks into a cleft formed by two ATPase domains which come together in an inactive dimer (packing front-to-front rather than the front-to-back in the active state). The substantial contact area between the receiver and ATPase domains reinforces the dimerization of the receiver domain. Upon activation by phosphorylation the receiver domain is reorganized into an alternate stable dimer, and the crossed helical pair extending from it is lost together with the interface to the ATPase domain (29). With the interaction of the receiver domain removed by activation, the ATPase domains can reorganize to active state front-to-back packing, ready to interact with another dimer to assemble the full ring required for activation. As noted above, the NtrC1-DBD is essentially monomeric in solution on its own, interacting only weakly upon binding to DNA. However the attached domains are always dimers, which removes selective pressure for tight dimerization of the DNA binding domain in this case.

NtrC4 is closer to NtrC1 in regulatory mechanism than NtrC, but with some significant differences that have structural consequences. The receiver domains of NtrC4 form the same basic interface with one another as seen in NtrC1 but lack the helical extension that forms a significant part of the interface with the ATPase domain in that protein. The receiver domain of NtrC4, expressed as a separate domain, only weakly dimerizes (27). This unactivated dimer does bind to the front-to-front inactive ATPase dimer, and does repress assembly of the active, hexamer ATPase. This repression is, however, very weak compared that in NtrC1 and raising the protein concentration leads to activity even in the absence of activating phosphorylation. Upon phosphorylation the receiver domain changes conformation in the same way that NtrC1 does, and becomes a more stable dimer. This activated dimer packs less well with the inactive ATPase dimer, releasing it to form front-to-front dimers and higher oligomers. It is worth noting that the characterization of activity was done at 70° C, well above the temperature for structural work but below the optimum growth temperature for *Aquifex aeolicus*. Both structural work and supporting biophysical characterization have shown that the NtrC4 DBD has the buttressing A helix that substantially stabilizes dimer formation. Given the weak dimerization of the receiver domains, and presumably the ATPase domains in the inactive state, it is necessary for NtrC4 to have the DNA binding domains dimerize to have high affinity DNA binding in the inactive state.

The Nlh2 regulatory GAF domains have been shown to be stable dimers in their ligand-free state, and to have a helical extension that is analogous to that in NtrC1. SAXS data support a model in which the dimeric GAF domains stabilize an inactive dimer state of the ATPase domains (40). Although the activated state of Nhl2 has not been characterized, the very similar GAF domains from Nhl1 (also from *A. aeolicus*) were shown to switch dimerization interface, suggesting a 'release' mechanism analogous to NtrC1. GAF domains have been found as stable dimers in both unactivated and activated states, and through analogy with NtrC1, dimerization of the DNA binding domain is not needed to maintain the overall dimeric state. This is consistent with the observed monomer state of the Nhl2 DBD when studied as a separated domain. Modeling suggests that the two Nhl2 DBDs of the dimer will interact in a manner analogous to NtrC1 and ZraR when bound to DNA.

There is as yet little understanding of the regulation for ZraR. The full length protein has an N-terminal receiver domain, which was deleted in the single structure solved for this protein. The structure solved of the ATPase-DBD construct shows that it spontaneously assembles to

the presumably active ATPase hexamer. The small dimerization interface between the DBDs observed in the crystal, and that fact that density was seen only for some of the DNA binding domains in the crystalline state, suggests that the DBDs are not inherently very stable dimers. As such, extending from the cases discussed above, we would predict that the ZraR receiver domains form stable dimers in both the unactivated and activated states, as do those from NtrC1 and DctD. Analysis of the amino acid sequence does not indicate the presence of a helical extension which helps stabilize the off-state regulatory domain dimers of NtrC1 and DctD, and furthermore the knobs-in-holes dimerization motif for the activated state that is present in NtrC1 and DctD (29) is apparently not present. This leaves mechanism for regulation of ZraR unclear, and detailed analysis of the oligomerization of the ZraR receiver domain will require direct study.

CD linkers of σ^{54} activators are disordered

A number of different lines of evidence suggest that the linkers between ATPase and DNA binding domains in σ^{54} transcriptional activators are disordered. In sequence alignments even among the small set of proteins studied here, the length and sequence of the linker region separating the central ATPase and DNA binding domains are highly variable, Figure 3. These linkers are also relatively high in proline residues, making formation of helices unlikely. Secondary structure prediction (Jpred) (41) does not indicate any preferred secondary structure in the linker regions for *Aquifex aeolicus* activators or mesophilic NtrC. The construct used for the NMR structure of NtrC's DNA binding domain (22) contained the full linker region. The chemical shifts and relaxation behavior of the linker residues showed clearly that they are a very dynamic random coil (moving with much shorter correlation time than the folded DBD). In the crystal structure of NtrC4-DBD (42) the linker residues are disordered, not giving rise to any density. NMR data on this NtrC4 construct again showed that the linker residues are very flexible in solution. In the only crystal structure of a σ^{54} activator including both central and the DNA binding domains, that of ZraR, both ATPase hexamers and DNA binding domain dimers were observed, but no density was seen for any of the linkers between these domains (20).

Crystal structure of full-length NtrC1: the DBD is flexibly tethered to the C domain

No structures have been previously reported for full-length intact σ^{54} activators, possibly due to the difficulty crystallizing multidomain proteins with flexible linkers. In the structure of full length NtrC1 the receiver and ATPase domains are very similar to the previously solved structure of a construct containing just those two domains. The most interesting feature of the new structure is that there is almost no electron density visible in the region where the DNA-binding domains are expected. This observation indicates that the DNA-binding domain positions are disordered in the crystal. There is a gap between the receiver domain–ATPase domain dimers in the crystal that is large enough to accommodate the DNA-binding domains. This gap separates the ATPase domains of one dimer from the receiver domains in the neighboring unit cell (Figure 5). The vertical distance between the central domains of one NtrC1 dimer and the receiver domains of in the next unit cell leaves sufficient room for the DNA binding domain, but it is a fairly tight fit, constraining the DBD's vertical position, however the DBD is less constrained in the horizontal dimension. The sparse lateral contacts enable the DNA-binding domain to have some rotational freedom around the long axis of the dimer as well as some translational freedom (Figure 4), either of which would explain the lack of electron density.

This structure provides additional evidence that the CD-linker in σ^{54} -proteins acts as an unstructured, flexible tether. In spite of a very small volume in the crystal for the DNA binding domain to occupy, and abundant surrounding surfaces to interact with, only very weak, disconnected density was observed in the crystals of full length NtrC1. If linkers

interacted with the ATPase domain, with one another, or with other parts of the DNA binding domain we would have anticipated observing some well-ordered density. However, in light of the data from other activators, it seems likely that the CD linker just serves as a flexible tether.

The single case among σ^{54} activators in which there is any evidence for localization of the DBDs is in assembled, hexameric NtrC studied by EM and SAXS. When the rings had ADP bound there was no clear density for the DBDs, however when an ATP analog was bound (ADP-AIF_x) there were regions of density under the ATPase ring (DeCarlo 2006). Those studies were at low resolution, so it is not clear how well localized the DBDs are (i.e. whether they would have been seen in a crystal structure). However, the localization could be due to surface interactions that occur between the oligomerized ATPase domains and the DBD, rather than being mediated by the ATPase to DBD linker, and the linkers could still be disordered.

Interdimer interactions and cooperative binding

It has been shown that dimers of transcriptional activators (including NtrC, DctD, NtrC1 and NtrC4) bind at tandem repeats of the specifically recognized DNA sequences. The binding at neighboring DNA sites occurs with positive cooperativity for NtrC in the unactivated state (9, 13, 43–44). On the other hand for NtrC1 the binding isotherms were well fitted by an independent binding model (29), though the experiments were done at temperatures well below the normal growth temperature for *Aquifex*. Upon activation, which enables assembly of hexameric rings of ATPase domains, the binding cooperativity is increased for NtrC as one would expect with the extensive interaction of those domains. Although NtrC ATPase domains cannot assemble without the aid of interactions through the receiver domains (11), it is possible that there are sufficient inter-ATPase domain interactions to lead to the observed cooperativity in binding even in the absence of phosphorylated receiver domains. For NtrC1 the receiver and ATPase domains dock tightly in the unactivated state to repress assembly of ATPase rings (10). In this case the surfaces of the ATPase domain used for assembly are not available for inter-dimer interaction. Models of two NtrC1 dimers bound at neighboring sites (Figure 5) show that they could interact but they apparently do only weakly if at all.

Flexible linking of DNA binding domains allows ATPase domain rearrangement

In previous studies of transcriptional activator regulatory mechanisms and activation, there has been little discussion of the DNA binding domains and their linkers to the ATPase domain. For all of those characterized there is a substantial reorganization of both the regulatory and ATPase domains that must occur while the activator must remain bound to the enhancer DNA region. The swiveling of the ATPase from an inactive front-to-front configuration into the active, front-to-back backed active ring configuration changes the relative positions of the end of the ATPase domain where the linker to the DNA binding domain connects (Figure 6). For this to occur without perturbation of the DNA-bound DBD it makes sense for it to be flexibly linked.

In other protein systems, there are numerous examples of linkers that act as simple tethers including immunoglobulins (45), leucyl-tRNA synthetase (46), xylanase Cex (47), as structured mediators of docking interactions or cellular cross-talk, Smad3 (48), DNA Pol (49), poly(rC) binding protein 2 (50), or as regions that can interconvert between structured and unstructured states in order to mediate docking/binding interactions, lac repressor (51, 52), retinoblastoma protein (53).

The nature of the dimer interface and DNA binding

As described in previous sections, NtrC, NtrC4 and Fis form stable dimer interfaces, while NtrC1, Nhl2 and ZraR are only marginally stable. Stability correlates strongly with the presence of the N-terminal A helix in these NtrC homologs. In other contexts a great diversity of N-terminal or C-terminal helices have been found in conjunction with the helix-turn-helix motifs, serving a variety of functions including forming additional dimerization interfaces, Spo0A (54), providing contacts to other domains or coordinating substrates, DosR (55), and communicating conformational changes to the DNA-binding mode, LacI (56).

Although it appears stable dimerization of the DBD is associated with need for overall dimerization, and therefore with the nature of the assembly control in the regulatory and ATPase domains, it is possible that dimerization of the DBD also may affect its DNA interactions. In the case of Fis, the extended dimerization interface formed by the long A helix stabilizes the dimer, and enables it to bend DNA (57, 58). Like Fis, NtrC's DNA-binding domain also has an A helix, and the spacing of its recognition helices suggests that it also bends DNA (22). There is also experimental evidence showing that NtrC can activate from bent DNAs that do not contain consensus binding sites, suggesting that the binding to bent DNA is high affinity. This further supports the idea that the A helix, or some other form of dimer stabilization, may be required for bending. The NtrC4-DBD, which has an abbreviated A helix encompassing 5 residues (compared to 13 in NtrC and 15 in Fis), was shown to cause a very slight bend in the DNA upon binding (28).

In the cases of NtrC1 and Nhl2 DNA-binding domains, secondary structure prediction, NMR, and mass spectrometry data suggest that the N-terminal region does not contain an A helix. The structures of the ZraR-DBD and the NtrC1-DBD overlay closely with an RMSD of 2.2 Å, have similar spacing between the recognition helices (ZraR: 16 Å, 1DBD: ~18Å), and bury similar solvent-accessible surface areas (ZraR: ~1070Å², 1DBD: ~1050Å²). One notable difference between NtrC1-DBD and ZraR is that ZraR has two hydrophobic contacts between the B helices (V405 and V409), where NtrC1-DBD has one primary hydrophobic contact (I399). In this subset of the activators it is possible that the hydrophobic contact between the B helices (e.g. Ile399 in NtrC1) constitutes a pivot point, allowing flexibility in the dimerization interface that enables the recognition helices to swing outward. This kind of pivot would presumably remove the ability of the protein to bend DNA as the protein structure would likely adapt to the structure of the recognized duplex DNA rather than the DNA to the protein. Pivots with functional purposes have been documented in helix-turn-helix DNA-binding proteins in the cases of the LacI headpiece (55) and the phage-encoded activator protein Mor (59), where the pivot enables the protein to switch between specific and nonspecific interaction modes, or recognize multiple binding sites. Further DNA complex structures are needed to extend understanding of this phenomenon.

Summary

With the set of transcriptional activator structures now available we can see that the dimerization determinant is variable and depends on the details of the activation mechanism. The DNA binding domains all contain conventional helix-turn-helix motifs, but the way they are presented for DNA recognition is somewhat variable. It appears that these are sometimes rigid units when strongly dimerized but may be flexible when the dimerization interface is small. As is often the case evolutionary adaptation in one domain may be followed by complementary changes in another tethered domain (60).

MATERIALS AND METHODS

Constructs

Secondary structure predictions suggest that the poorly-conserved linker regions between the conserved C and D domains in NtrC1, Nlh2, and NtrC4 are unstructured. Constructs were selected starting from a point near the start of the predicted flexible linker and ending at the C-terminus of the full-length protein. Two different constructs of NtrC1DBD were made, NtrC1DBD(373–439) (“1DBD-long”), including 20 residues of the N-terminal flexible linker, and NtrC1DBD(392–439) (“1DBD-short”) that completely removes the N-terminal linker region. The construct of Nlh2 includes residues 428–497 (2DBD). The NtrC4 DNA binding domain construct with residues 365–442 (4DBD), was described previously (Batchelor 2008), and contains most of the linker region.

Protein Expression and Purification

1DBD (residues 392–439), 2DBD (residues 365–442), and 4DBD (residues 428–497) were separately cloned into the pET21a vector (1DBD, 4DBD) and the pET21b vector (2DBD) as a (His)₆tag-fusion protein. Cells were transformed into *E. coli* BL21(DE3) with a Rosetta.pLysS plasmid, and grown to an OD₆₀₀ of 0.8 (1DBD, 4DBD) or 0.5 (2DBD) in Luria Broth media. Cells were induced with 1 mM isopropyl β-D-thiogalactopyranoside and harvested after 12–16 hours at 25°C (1DBD) and 4–6 hours at 37°C (2DBD). Isotopically labeled proteins were switched to ¹⁵N or ¹⁵N/¹³C-labeled M9 media (Cambridge Isotope Laboratory) before induction as described (61) and induced for 4–6 hours. Cells were harvested by centrifugation at 5000 rpm for 30 minutes. Cell pellets were suspended in lysis buffer (1DBD/2DBD: 50 mM sodium phosphate pH 6.5, 200 mM NaCl, 1 mM EDTA, and 0.1 mM PMSF; 2DBD: 20 mM sodium phosphate pH 7.0, 0.5 M NaCl, 0.1 mM PMSF), and disrupted by sonication. 1DBD was additionally heated to 80°C for 30 minutes to precipitate *E. coli* proteins. Lysates were then centrifuged at 30,000g for 30min, and supernatants were filtered using 0.45 μm filters to remove insoluble material. 1DBD was purified by heparin affinity chromatography and eluted with a continuous linear NaCl gradient (200 mM–700 mM). 2DBD was loaded onto a HiTrap™ Chelation HP (Pharmacia Biotech.) column, and bound protein was eluted with a continuous 100 mL linear gradient of imidazole (0 – 500 mM). The His-tag was cleaved from 2DBD using tobacco etch virus (TEV) protease at 30°C for 6 hours. The cleaved His₆-tag and uncleaved protein were removed by loading the digested proteins over the HiTrap™ Chelation HP again. 1DBD, 2DBD, and 4DBD were finally purified by gel filtration on a HiLoad Superdex 75 prep grade column (Pharmacia) in gel filtration buffer (1DBD and 4DBD: 50 mM sodium phosphate buffer pH 6.5, 100 mM NaCl, 1 mM EDTA; 2DBD: 20 mM sodium phosphate pH 7.0, 100 mM NaCl). Samples were concentrated using Amicon Ultra centrifugal concentrators (5 kDa cutoff; Millipore).

NMR spectroscopy

NMR experiments were conducted on DRX-Bruker 800 MHz, 600 MHz, and 500 MHz spectrometers with triple resonance cryoprobes and a triple-axis gradient amplifiers. Experiments were conducted at 303K. The following experiments were used to make assignments for backbone and side-chain atoms: ¹⁵N-HSQC, ¹³C-HSQC, HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, C(CCCO)NH, H(CCCO)NH, HBHA(CO)NH and HCCH-TOCSY experiments. For NOE assignments, ¹⁵N-edited NOESY-HSQC and ¹³C-edited NOESY-HSQC spectra in water and D₂O were used. Spectra were processed with the NMRPipe/nmrDraw software package and analyzed using the program Sparky. Backbone dihedral angle constraints were derived from ¹H_α, ¹³C_α, ¹³C_β chemical shifts with the program TALOS. Hydrogen bond constraints were determined based on the slow exchanging amide protons identified by recording ¹H-¹⁵N HSQC spectra after an ¹⁵N-labeled protein sample had been exchanged into a D₂O buffer. ¹H chemical shifts were referenced

to the methyl proton signals of DSS of 0 ppm, and ^{13}C and ^{15}N chemical shifts calculated indirectly from the ^1H spectrometer frequency.

Protein crystalization and structure determination

Protein crystallization was done using standard hanging drops using 96 well plates. The NtrC1 DBD crystallized at 5 mg/ml from 25% PEG 4000, 0.2 M ammonium sulfate, 0.1 M sodium acetate, pH 4.6. Before freezing PEG was increased to 35%. Crystals formed as clusters of plates, data were collected using the end of one well separated plate. NtrC1 RCD crystallized from 2.5 M sodium chloride, pH 7.5, giving rod-like crystals. Data were collected on beamline 8.3.1 at the Advanced Light Source, Lawrence Berkeley Lab. Data were processed and scaled with MOSFLM (62) and the CCP4 program suite (63) using the ELVES interface (64). Molecular replacement used the PHASER implementation of phenix.automr (65), with the NtrC4 DNA binding domain (3E7L) or the NtrC1 receiver and central domain construct (1NY5) as search model for NtrC1-DBA and NtrC1-RCD respectively. Phenix.autobuild was used to generate an initial model of NtrC1-RCD. Amino acid sequences were fitted to the density using Coot (66). Refinements were done using with phenix.refine (67). Structures have been deposited to the PDB under ID codes 4L5E (for NtrC1 DBD) and 4L4U (for full length NtrC1).

NMR Structure calculations

A structure calculation for NtrC1-DBD was done using a semi-automated protocol with CYANA 2.1. The initial fold was obtained by manual NOE assignment and peak picking for secondary topology. NOE cross-peak assignments were obtained by using manual procedures. An initial fold of the protein was calculated on the basis of manually assigned NOEs, essential for secondary topology determination, with subsequent use of the program CYANA. Seven cycles of CYANA routine were calculated and each cycle had 10000 steps of torsion-angle dynamics with a simulated annealing protocol. 100 structures were calculated and 20 structures with the lowest target function values were chosen for analysis. A summary of NMR-derived restraints and structures are shown in Table 2. The 20 final structures with lowest NOE energies were retained and validated by the program PROCHECK. Structures were analyzed and visualized using the program PyMOL, and MOLMOL. The family of 20 structures has been deposited to the PDB under ID 2M8G.

Analytical ultracentrifugation

Analytical ultracentrifugation experiments with 1DBD-long, 2DBD, and 4DBD were carried out on a Beckman ultracentrifuge model XL-I using a 12-mm six-channel centerpiece and an An-50 Ti rotor. Data were collected at 20000 rpm at 280 nm and 300 nm for protein concentrations of 0.5, 1.0 and 1.4 mg/ml in 250 mM NaCl, 50 mM NaPO₄ pH 6.5. Each run was allowed to reach equilibrium by spinning the samples for 24 hours. Absorbance scans were taken in 10 replicates for each sample at 20°C. The radial step size was 0.001. Data analysis was performed using the software UltraScan-7.0.

Structural comparisons

Structures were visualized with PYMOL. Helix angles, distances, and buried surface area were calculated with CHIMERA. In order to describe the widths of the minor groove and the spacing between the recognition helices consistently, the following system was used: in DNA, measurements were taken from the shortest length between two distally-separated O atoms in the phosphate backbone of DNA, and in protein, measurements were taken from the shortest distance between two residues in opposing α helices, including the side chain only to the C β carbon, to allow for side chain flexibility.

Acknowledgments

We thank Dr. Gongyi Zhang for providing initial diffraction data on full length NtrC1, Prof. Sydney Kustu for helpful discussions, and Alex Siegel for helping with figures. This work was supported by the National Institutes of Health (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.

References

1. Burgess RR, Travers AA, Dunn JJ, Bautz EK. *Nature*. 1969; 221:43–46. [PubMed: 4882047]
2. Burgess RR. *J Biol Chem*. 1969; 244:6168–76. [PubMed: 4900511]
3. Wosten MM. *FEMS Microbiol Rev*. 1998; 22:127–50. [PubMed: 9818380]
4. Wignestweraraj S, Bose DM, Burrows PC, Joly N, Schumacher J, Rappas M, Pape T, Zhang X, Stockley P, Severinov K, Buck M. *Mol Microbiol*. 2008; 68:538–46. [PubMed: 18331472]
5. Kazmierczak MK, Wiedmann M, Boor KJ. *Mol Biol Rev*. 2005; 69:527–43.
6. Buck M, Gallegos MT, Studholme DJ, Guo Y, Gralla JD. *J Bact*. 2000; 182:4129–36. [PubMed: 10894718]
7. Reitzer L, Schneider BL. *Microbiol Mol Biol Rev*. 2001; 65:422–33. [PubMed: 11528004]
8. Model P, Jovanovic G, Dworkin J. *Mol Microbiol*. 1997; 24:255–61. [PubMed: 9159513]
9. Porter SC, North AK, Kustu S. *Genes & Dev*. 1993; 7:2258–73. [PubMed: 7901122]
10. Lee SY, De La Torre A, Yan D, Kustu S, Nixon BT, Wemmer DE. *Genes & Dev*. 2003; 17:2552–63. [PubMed: 14561776]
11. De Carlo S, Chen B, Hoover TR, Kondrashkina E, Nogales E, Nixon BT. *Genes & Dev*. 2006; 20:1485–1495. [PubMed: 16751184]
12. Studholme DJ, Dixon R. *J Bact*. 2003; 185:1757–67. [PubMed: 12618438]
13. Reitzer LJ, Magasanik B. *J Bact*. 1989; 171:5512–22. [PubMed: 2571609]
14. Su W, Porter S, Kustu S, Echols H. *Proc Natl Acad Sci*. 1990; 87:5504–8. [PubMed: 2164685]
15. Wedel A, Weiss DS, Popham D, Droge P, Kustu S. *Science*. 1990; 248:486–90. [PubMed: 1970441]
16. Zhang X, Chaney M, Wigneshweraraj SR, Schumacher J, Bordes P, Cannon W, Buck M. *Mol Microbiol*. 2002; 45:895–903. [PubMed: 12180911]
17. Hastings CA, Lee SY, Cho HS, Yan D, Kustu S, Wemmer DE. *Biochem*. 2003; 42:9081–90. [PubMed: 12885241]
18. Meyer M, Park S, Zeringue L, Staley M, McKinstry M, Kaufman RI, Zhang H, Yan D, Yennawar N, Yennawar H, Farber GK, Nixon BT. *FASEB J*. 2001; 15:1326–28. [PubMed: 11344129]
19. Rappas M, Schumacher J, Beuron F, Niwa H, Bordes P, Wigneshweraraj S, Keetch CA, Robinson CV, Buck M, Zhang X. *Science*. 2005; 307:1972–75. [PubMed: 15790859]
20. Sallai L, Tucker P. *J Struct Biol*. 2005; 151:160–170. [PubMed: 16005641]
21. Schumacher J, Joly N, Rappas M, Zhang X, Buck M. *J Struct Biol*. 2006; 156:190–9. [PubMed: 16531068]
22. Pelton JG, Kustu S, Wemmer DE. *J Mol Biol*. 1999; 292:1095–1110. [PubMed: 10512705]
23. Hunt TP, Magasanik B. *Proc Natl Acad Sci*. 1985; 82:8453–57. [PubMed: 2867543]
24. Hirschman J, Wong PK, Sei K, Keener J, Kustu S. *Proc Natl Acad Sci*. 1985; 82:7525–29. [PubMed: 2999766]
25. Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, Graham DE, Overbeek R, Snead MA, Keller M, Aujay M, Huber R, Feldman RA, Short JM, Olson GJ, Swanson RV. *Nature*. 1998; 392:353–58. [PubMed: 9537320]
26. Martinez-Argudo I, Little R, Shearer N, Johnson P, Dixon R. *J Bacteriol*. 2004; 186:601–10. [PubMed: 14729684]

27. Batchelor JD, Doucleff M, Lee CJ, Matsubara K, De Carlo S, Heideker J, Lamers MH, Pelton JG, Wemmer DE. *J Mol Biol.* 2008; 384:1058–75. [PubMed: 18955063]
28. Vidangos N, Heideker J, Lyuvimov A, Lamers M, Huo Y, Pelton JG, Ton J, Gralla J, Berger J, Wemmer DE. submitted.
29. Doucleff M, Chen B, Maris AE, Wemmer DE, Kondrashkina E, Nixon BT. *J Mol Biol.* 2005; 353:242–55. [PubMed: 16169010]
30. Dominguez C, Boelens R, Bonvin AJ. *J Am Chem Soc.* 2003; 125:1731–37. [PubMed: 12580598]
31. de Vries SJ, van Dijk ADJ, Krzeminski M, van Dijk M, Thureau A, Hsu V, Wassenaar T, Bonvin AMJJ. *Proteins: Struct Funct & Bioinf.* 2007; 69:726–733.
32. North A, Kustu S. *J Mol Biol.* 1997; 267:17–36. [PubMed: 9096204]
33. Schulz A, Langowski J, Rippe K. *J Mol Biol.* 2000; 300:709–725. [PubMed: 10891265]
34. Morett E, Bork P. *FEBS Lett.* 1998; 433:108–12. [PubMed: 9738943]
35. Aravind L, Anantharaman V, Balaji S, Babu MM, Lakshminarayan MI. *FEMS Microbiol Rev.* 2005; 29:231–62. [PubMed: 15808743]
36. Rohs R, Jin X, West SM, Joshi R, Honig B, Mann RS. *Ann Rev Biochem.* 2010; 79:233–69. [PubMed: 20334529]
37. Huffman JL, Brennan RG. *Curr Opin Struct Biol.* 2002; 12:98–106. [PubMed: 11839496]
38. Chen P, Reitzer L. *J Bact.* 1995; 177:2490–96. [PubMed: 7730282]
39. Klose KE, North AK, Stedman KM, Kustu S. *J Mol Biol.* 1994; 241:233–45. [PubMed: 8057363]
40. Batchelor JD, Lee PS, Wang AC, Doucleff M, Wemmer DE. *J Mol Biol.* 2013; 425:156–70. [PubMed: 23123379]
41. Cole C, Barber JD, Barton GJ. *Nuc Acids Res.* 2008; 36:W197–201.
42. Vidangos, N. PhD Thesis. University of California; Berkeley: 2010.
43. Weiss V, Claverie-Martin F, Magasanik B. *Proc Natl Acad Sci.* 1992; 89:5088–92. [PubMed: 1350679]
44. Scholl D, Nixon BT. *J Biol Chem.* 1996; 271:26435–26442. [PubMed: 8824302]
45. Ababou A, Rostkova E, Mistry S, Le Masurier C, Gautel M, Pfuhl M. *J Mol Biol.* 2008; 384:615–30. [PubMed: 18926831]
46. Hsu JL, Martinis SA. *J Mol Biol.* 2007; 376:482–91. [PubMed: 18155724]
47. Poon DK, Wither SG, McIntosh LP. *J Biol Chem.* 2007; 282:2091–100. [PubMed: 17121820]
48. Vasilaki E, Siderakis M, Papakosta P, Skourti-Stathaki K, Mavridou S, Kardassis D. *Biochem.* 2009; 48:8366–78. [PubMed: 19645436]
49. Klinge S, Núñez-Ramírez R, Llorca O, Pellegrini L. *EMBO J.* 2009; 28:1978–87. [PubMed: 19494830]
50. Sean P, Nguyen JH, Semler BL. *Virology.* 2008; 378:243–53. [PubMed: 18656221]
51. Kercher MA, Lu P, Lewis M. *Curr Opin Struct Biol.* 1997; 7:76–85. [PubMed: 9032054]
52. Swint-Kruse L, Matthews KS. *Curr Opin Microbiol.* 2009; 12:129–37. [PubMed: 19269243]
53. Burke JR, Deshong AJ, Pelton JG, Rubin SM. *J Biol Chem.* 2010; 285:16286–93. [PubMed: 20223825]
54. Lewis RJ, Scott DJ, Brannigan JA, Ladds JC, Cervin MA, Spiegelman GB, Hoggett JG, Barak I, Wilkinson AJ. *J Mol Biol.* 2002; 316:235–45. [PubMed: 11851334]
55. Wisedchaisri G, Wu M, Sherman DR, Hol WGJ. *J Mol Biol.* 2008; 378:227–242. [PubMed: 18353359]
56. Kalodimos CG, Folkers HE, Boelens R, Kaptein R. *Proc Natl Acad Sci.* 2001; 22:6039–44. [PubMed: 11353825]
57. Cheng Y, Wie-Zen Y, Johnson R, Yuan H. *J Mol Biol.* 2000; 302:1139–51. [PubMed: 11183780]
58. Stella S, Cascio D, Johnson RC. *Genes & Dev.* 2010; 24:814–26. [PubMed: 20395367]
59. Kumaraswami M, Howe MM, Park HW. *J Biol Chem.* 2004; 279:16581–90. [PubMed: 14729670]
60. Kuriyan J, Eisenberg D. *Nature.* 2007; 450:983–90. [PubMed: 18075577]
61. Marley J, Lu M, Bracken C. *J Biomol NMR.* 2001; 20:71–75. [PubMed: 11430757]
62. Powell H. *Acta Cryst D Biol Cryst.* 1999; 55:1690–95.

63. Bailey S. *Acta Cryst D Biol Cryst.* 1994; 50:760–63.
64. Holton J, Alber T. *Proc Natl Acad Sci.* 2004; 101:1537–42. [PubMed: 14752198]
65. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storone LC, Read RJ. *J Appl Cryst.* 2007; 40:658–74. [PubMed: 19461840]
66. Emsley P, Cowtan K. *Acta Cryst D Biol Cryst.* 2004; 60:2126–32.
67. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Twerwilliger TC. *Acta Cryst D Biol Cryst.* 2002; 58:1948–54.
68. Santero E, Hoover TR, North AK, Berger DK, Porter SC, Kustu S. *J Mol Biol.* 1992; 227:602–20. [PubMed: 1404379]

Figure 1. Schematic diagrams of transcription initiation at σ^{70} (left) and σ^{54} (right) polymerase complexes at promoters. For σ^{54} mediated initiation the protein IHF can help bend DNA to facilitate contact of the activator and polymerase (68).

Central Domain						
NtrC1	IEIPPLRERKEDIIP	LANHFLKKFSRKYAKEVEGFTKSAQ	ELLLSY	PWYGNVRELKNVIE	364	
ZraR	IEMPSLRQRREDIPL	LADHFLRRFAERNRKVVKGFT	PQAMD	LLIHWDWPGNIRELENAIE	366	
Nlh2	LYVPPLRERKEDIPI	LIEYYL	KELNKKYGKEYV	-LSEDALEVLLEYDYPGNIRELINILE	417	
NtrC4	IYLPPLRERKGDVILL	AEYFLKKFAKEYKKNCFEL	SEETKEYLMKQEWK	GNVRELKNLIE	356	
NtrC	IHLPLRERREDIPRL	ARHFLQVAARELGVEAKLLHPETETAL	TRLAWPGNVRQLENTCR		365	
	C	C-D linker	Helix A			
NtrC1	RAVLFS	SEGKFI	DRGELSC	LVNS-----KG	IKN-----	391
ZraR	RAVVLT	GGEYISERELP--	LATAATP-----	IKTEY	SG-----	397
Nlh2	RVVILNNG-RV	TSSELPELLRKR	-----ERKTG-----	D		445
NtrC4	RAVILCEGE	VIKPEDLGLKEKSWRD	-----LSYLL	IKI-----		390
NtrC	WLTVMAGOEVLIC	DLPGELEASTPDSPSHLPPDS	WATLLAQWADRALLR	SGHQN-L		421
Fis	MFE-QRVNS-DVLT	VSTVNSQDQVTQKPLRDSV	KQALLANYFAQL	NGQDVND		
DNA Binding Domain						
		Helix B	Helix C	Helix D		
NtrC1		KHKS	IKIEKEEIIKVLKEVNFNKKLASETI	GIP	LRTLYRRLKEYGIE	439
ZraR		EIQPLVDVEKEVILAALKETGGNKTEAARQI	GIT	RKTLLAKISR		441
Nlh2		LPKFIEETEKKRIIEALEFTGYVKSRAAKLI	GYT	LRQLDYRIKKYGIELKKF		497
NtrC4		LKEAKKEFEKIFIEEKLREYDYDLKRTAEI	GID	LSNLYRKIKSINIRVKSS		442
NtrC		LSEAQPELERTLLTALRHTQGHKQEAARI	LGWGR	NTLTKKLKELGME		469
Fis		VELVLAEVEQPLDMVMQYTRGNQTRAALM	MGIN	RGTLRKKLKKYGMN		

Figure 2.

Sequence alignment of the C-terminal and linker regions of transcriptional activators. NtrC, ZraR sequences are from *S. enterica* serovar *Typhimurium*, FIS is from *E. coli*, and NtrC1, Nlh2 and NtrC4 are from *A. aeolicus*. Medium grey regions mark experimentally determined helices, light grey regions have not had secondary structure determined. The highlighted color overlay indicates amino acids in constructs of DBDs studied structurally. For NtrC1 the vertical line indicates the beginning of the short (linkerless) construct studied.

Figure 3. Structures of the NtrC DNA-binding domains. a) Helix nomenclature in helix-turn-helix dimers. b) Overlay of monomeric subunits of NtrC (purple), NtrC1 (green), Nlh2 (teal), and NtrC4 (yellow) FIS (dark red) DBDs.

Figure 4.
Comparison of dimers of DNA-binding domains from *A. aeolicus* and other orthologues.

Figure 5.

a) Model for placing NtrC1-DBD into full-length NtrC1 crystal structure. b) A model of two dimeric NtrC1-full length activators on Aq_1119 native DNA tandem binding sites showing that the receiver and ATPase domains are in close proximity.

Figure 6.

Models are shown for the ‘on’ state (left) and ‘off’ state of the NtrC4 transcriptional activator showing receiver domains (top), ATPase domains (center) and DNA-bound DBDs (bottom). In the off state the receiver domains dock against the ATPase domains, holding them in a front-to-front configuration. In this configuration the C-termini of the ATPase domains (red spheres) are relatively close together. In the activated, ‘on’ configuration the receiver domains and ATPase domains change their dimer interfaces. This positions the C-termini of the ATPase domains significantly further apart. Flexible linkage between the ATPase and DNA binding domains is needed to accommodate this rearrangement.

Table 1

Summary of activator DNA binding domain structures.

Characteristic	NtrC1-DBD	NtrC2-DBD	NtrC4-DBD	NtrC-DBD
Organism	<i>A. aeolicus</i>	<i>A. aeolicus</i>	<i>A. aeolicus</i>	<i>S. typhimurium</i>
construct	392–439	428–497	374–442	
Structure solved	Crystallography	NMR	Crystallography	NMR
Resolution or r.m.s.d.	1.3 Å	0.7 Å r.m.s.d.	2.2 Å	1.2 Å r.m.s.d.
AUC oligomer state	monomer	monomer	dimer	dimer
“A helix”	unstructured	unstructured	5.6–7.0 Å	31.2 Å
B-helix angle	33°	(≈20°)	42°	14°
residues in recognition helix *	<u>L</u> R <u>T</u> L <u>Y</u> R <u>R</u> L <u>K</u> E <u>Y</u>	<u>L</u> R <u>Q</u> L <u>D</u> Y <u>R</u> I <u>K</u> K <u>Y</u>	<u>L</u> S <u>N</u> L <u>Y</u> R <u>K</u> I <u>K</u> S <u>L</u> N	<u>L</u> K <u>E</u> A <u>K</u> K <u>E</u> F <u>E</u> K <u>I</u> F <u>I</u> E <u>E</u> K
dimerization interface **	<u>S</u> I <u>K</u> E <u>I</u> E <u>K</u> E <u>E</u> I	<u>P</u> K <u>F</u> I <u>E</u> E <u>T</u> E <u>K</u> K <u>R</u>	<u>L</u> K <u>E</u> A <u>K</u> K <u>E</u> F <u>E</u> K <u>I</u> F <u>I</u> E <u>E</u> K	<u>L</u> S <u>E</u> A <u>Q</u> P <u>E</u> L <u>E</u> R <u>T</u> L
Min. distance between central axes of D helices	22 Å	27 Å	28 Å	18 Å
angles formed by recognition helices	20°	52°	46°	37°
Accessible surface area monomer/dimer	3830/6600	3870/6660	6,370/10,350	8,750/14,900
Total buried SA (Å ²)	1050	1080	2,390	2,600

* Residues that are underlined are thought to be involved in interactions with DNA

** Residues that are underlined are involved in the hydrophobic packing of the dimerization interface Grey text refers to information inferred from HADDOCK models.

Table 2

Structure statistics.

	NtrC1-DBD	NtrC1 full length
Data Collection		
Space group:	C222 ₁	C222 ₁
Unit Cell parameters		
a, b, c	33.38, 62.27, 54.51	102.42, 107.01, 98.18
α , β , γ	90, 90, 90	90, 90, 90
Resolution (Å)	54.2–1.34 (1.39–1.34)	19.8–2.2 (2.28–2.2)
Unique Reflections	13002 (1340)	27022 (2890)
Completeness (%)	99.1 (93)	97.6 (95)
Refinement:		
R _{work} /R _{free} (%)	17.3/20.6	21.3/24.9
Number of atoms		
Protein	409	3119
Solvent	82	158
Other (SO ₄ ²⁻)	10	0
B-factors		
Protein	15.9	47.9
Water	34.0	44.1
RMSDs		
bond lengths (Å)	0.005	0.003
bond angles (°)	0.865	0.696

Values in parentheses correspond to the highest resolution shell.

[Structures deposited PDB ID 4L5E (NtrC1-DBD) and 4L4U (NtrC1 full length).

Table 3

Structural statistics for the final calculated structures

NOE distance restraints	number
All	1232
Intraresidue	344
Sequential ($ i-j = 1$)	306
Medium range ($2 < i-j \leq 5$)	309
Long range ($ i-j > 5$)	273
Hydrogen bonds distance restraints ^a (number)	21
Backbone dihedral angle	
All	86
Phi	43
Psi	43
Mean CYANA target function	0.39 ± 0.05
Mean RMS deviations from the average coordinate ^b	
Backbone atoms (N, C _{α} , C', O)	0.14 ± 0.05 Å
Heavy atoms	0.72 ± 0.07 Å
Ramachandran plot (%)	
Most favored regions	87.3%
Additional allowed regions	12.5%
Generously allowed regions	0.0%
Disallowed regions	0.2%

^aTwo restraints per one hydrogen bond.^brmsd values for residues 448–492

[structure deposited PDB ID 2M8G]