Published in final edited form as:

J Mol Biol. 2009 July 3; 390(1): 70–82. doi:10.1016/j.jmb.2009.04.070.

Structure of the RNA-Polymerase Core Binding Domain of σ^{54} Reveals a Likely Conformational Fracture Point

Eunmi Hong^{1,2}, Michaelleen Doucleff^{1,2}, and David E. Wemmer^{1,2}

Eunmi Hong: turtulee@spin.yonsei.ac.kr; Michaelleen Doucleff: doucleffm@niddk.nih.gov; David E. Wemmer: dewemmer@lbl.gov

¹Dept. of Chemistry, University of California, Berkeley, CA, 94720-1460, U.S.A

²Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, U.S.A

Summary

Transcription initiation by bacterial σ^{54} -RNA polymerase requires a conformational change of the holo-polymerase-DNA complex, driven by an enhancer binding protein. Although structures of the core polymerase and the more common σ^{70} factor have been determined, little is known about the structure of the σ^{54} variant. We report here the structure of an *Aquifex aeolicus* σ^{54} domain, residues 69–198, which binds core RNA polymerase. The structure is comprised of two distinct subdomains held together by a small, conserved hydrophobic interface that appears to act as a fracture point in the structure. The N-terminal four-helical subdomain has a negative surface and conserved residues that likely contact the core polymerase, while the C-terminal three-helical bundle has a strongly positive patch that could contact DNA. Sequence conservation indicates that these structural features are conserved and are important for the role of σ^{54} in the polymerase complex.

Keywords

sigma-54; bacterial transcription; RNA polymerase; NMR structure; gene regulation

Introduction

The cellular enzyme responsible for transcription is termed core RNA polymerase (RNAP)¹. It is comprised of five subunits ($\alpha_2\beta\beta'\omega$) that are fully capable of catalyzing the synthesis of RNA using a DNA template. However, transcription initiation at specific promoters requires additional proteins. In eukaryotic organisms, dozens of activators and accessory proteins are necessary for polymerase to bind the promoter and initiate transcription². In contrast, gene-specific transcription initiation in bacteria is relatively simple; it usually requires only one additional protein, called the σ factor^{3; 4}.

The σ factor acts by associating with the core RNAP to form the σ -RNAP holoenzyme. This complex binds to specific DNA sequences immediately upstream of a targeted gene, with the specificity provided solely by the σ factor. The σ factor ‘melts’ the double-stranded DNA to create a short region of single-stranded DNA, which serves as the initiation template for RNA synthesis⁵.

ACCESSION NUMBERS: Coordinates have been deposited in the Protein Data Bank with accession numbers: 2K9L for CBD(60-135) and as 2K9M for CBD(69-198) and NMR assignments have been deposited into the BMRB (accession numbers 15990 and 15991).

Studies of factors that regulate gene expression lead to the identification of two distinct classes of bacterial σ -factors, which are grouped by functional and sequence similarity with the first recognized members: σ^{70} and σ^{54} . The σ^{70} -class includes the major σ -factor, called σ^{70} in *Escherichia coli* (or σ^A in Gram-positive bacteria), which is responsible for transcription of most genes during exponential cell growth. Target genes are recognized by interactions of the σ -factor with consensus sequences at -35 base pairs relative to the start codon of the gene. Upon binding to promoter DNA, σ^{70} -RNAP holoenzyme can immediately open double-stranded DNA and spontaneously initiate transcription. Regulation of σ^{70} -dependent genes occurs primarily by modulating the amounts of free σ^{70} -type proteins available to bind RNAP⁶.

The second class of σ factor is uniquely represented by σ^{54} , for which no alternate forms have been identified. σ^{54} proteins have no detectable sequence homology to the σ^{70} -type factors and, although σ^{54} -factors associate with the same core RNAP, σ^{54} regulates transcription by a significantly different mechanism than σ^{70} ⁴. When σ^{54} -RNAP holoenzyme binds to promoter DNA, primarily through a consensus sequence at -24 base pairs relative to the start codon, the DNA in the complex remains in an inactive, closed state. To open the double-stranded promoter DNA and start transcription, σ^{54} -RNAP holoenzyme must interact with specialized enhancer binding proteins (EBPs). EBPs bind specific sequences relatively far upstream (typically 150 base pairs) from the genes they regulate. When activated by a cellular signal, the EBPs assemble into a catalytically competent oligomer and use the energy of ATP hydrolysis to remodel the σ^{54} -RNAP promoter complex. Only after this remodeling is polymerase in a conformation that is capable of opening the DNA and then initiating RNA synthesis⁷. Thus, rather than regulating the pool of available σ^{54} to bind RNAP, control of σ^{54} -dependent genes occurs primarily through generation of the cellular signals that enable assembly and activity of the EBPs, providing a fast response and high dynamic range of output.

Although often associated with their pivotal roles in transcription of genes involved in nitrogen metabolism in enteric bacteria (which lead to their initial discovery), σ^{54} and its EBPs are also essential for the virulence of many animal pathogens. For example, σ^{54} is required for the infection of *Borrelia burgdorferi*, the agent of Lyme disease⁸. In *Burkholderia cenocepacia*, the organism causing chronic, incurable respiratory infections in cystic fibrosis patients, σ^{54} is required for biofilm production and, possibly, for the pathogen's survival in human macrophages⁹. Thus, understanding σ^{54} and how it regulates gene transcription is of both intrinsic and practical interest.

The σ^{54} protein sequence has three functional regions (I–III)^{10; 11; 12}, Figure 1. Region I interacts directly with the EBPs and is thus required for coupling ATP hydrolysis with isomerization of the σ^{54} -RNAP-holoenzyme from the inactive to active states. Region II is a linker to Region III, which contains the core RNAP-binding domain and the DNA-binding domains (a segment that contacts the -12 region where promoter opening occurs, not yet characterized structurally, and a helix-turn-helix motif that recognizes the -24 promoter element¹³).

Although numerous biochemical and genetic studies have provided information about how these domains are important for transcription initiation¹⁴, high-resolution structures will facilitate understanding σ^{54} 's mechanistic details. σ^{70} factors have been extensively characterized by x-ray crystallography^{15; 16; 17; 18}, but σ^{54} has been recalcitrant to crystallization, and only low-resolution structures from SAXS and cryo-EM have been available^{19; 20}.

A divide-and-conquer approach yielded the first high-resolution structural information for σ^{54} . We recently showed that the last 60 residues of σ^{54} from *Aquifex aeolicus* form an independently-folded domain that specifically recognizes the major groove at the -24 region of promoter DNA¹³. However, this domain is only a small part of the 45-kDa protein (~400 amino acids). Further structural information is needed to understand how σ^{54} interacts with core RNAP, and σ^{54} 's role in promoter melting.

As the next step in addressing these questions, we present the solution structures of two σ^{54} fragments comprising the N-terminal part of Region III, which is essential for core RNAP binding²¹. The first construct, residues 60–135, folds into a highly stable four-helix bundle. A longer fragment with residues 69 to 198 (numbering used here corresponds to *A. aeolicus* unless stated otherwise) includes the first fragment plus 60 more residues at the C-terminal end, which fold into three α -helices that pack against the helical bundle of the smaller version (60–135). Previous mutagenesis studies, together with comparisons to σ^{70} structures, indicate that σ^{54} contacts core polymerase primarily via the four-helix bundle (60–135). Features of the structure also suggest a novel functional mechanism for this σ^{54} domain.

Results

Identification of structural domains in *A. aeolicus* σ^{54}

We prepared a recombinant protein encoding the region of *A. aeolicus* σ^{54} that, by sequence comparison with the well-characterized *E. coli* version, corresponded to regions that are primarily responsible for interaction with the core RNAP. This fragment (residues 48–183) was overexpressed in *E. coli* and purified by a fused N-terminal 6xHis tag. To identify potential structural subdomain boundaries in this region, samples were subjected to limited proteolysis at 310 K with a low concentration of proteinase-K. Two major fragments were detected on SDS-PAGE gels (data not shown). The molecular weights were determined to be 7.78 and 13.5 kDa by MALDI-TOF mass spectroscopy (data not shown). A single cleavage between residues 68 and 69 gave a very good match in predicted mass for the larger fragment, and so we cloned residues 69–183, and expressed and purified this as ¹⁵N-labeled protein. ¹H-¹⁵N HSQC spectra were collected at different temperatures and protein concentrations. Under the best conditions, only approximately 70% of the predicted number of cross peaks were visible in the ¹H-¹⁵N HSQC spectra with reasonable dispersion in chemical shifts. However, when we attempted to assign NMR resonances using triple resonance NMR experiments with a ¹³C- and ¹⁵N-labeled protein sample, many resonances could not be connected and assigned. These results strongly suggested that this σ^{54} fragment is partially unfolded.

Based on these results and secondary structure predictions, we designed new constructs of the RNAP-interaction region. Residue 60 was chosen as the N-terminus to prevent truncation of any secondary structure. In addition, four helices were predicted for residues 60–135, suggesting that this segment might be well-folded. Although there were two additional helices predicted at the C-terminal end of the 69–183 construct, the NMR spectra suggested that they were not fully folded, and hence they were deleted. The 60–135 construct gave excellent spectra (Figure 3, left), and its structure was determined. We then prepared a C-terminally extended construct that included residues 69–240. NMR spectra indicated that the C-terminal portion of this protein fragment is unstructured in this context, and so we made another version without the last 40 residues. This new construct (residues 69–198) includes the predicted four-helix domain in 60–135 plus three predicted helices at the C-terminal end. This protein fragment also yielded excellent spectra (Figure 3, right), and its structure was determined. Gel filtration elution times and NMR linewidths for all these fragments indicate that the proteins are monomeric in solution (data not shown).

To analyze whether the C-terminal extension, residues 138–240, is folded independently of the N-terminal four-helix bundle (60–135), these residues were also expressed as a separate construct. Although the fragment went into inclusion bodies during expression, the protein could be refolded and subsequently characterized by NMR.

Resonance assignment of σ^{54} 60–135, 69–198 and 69–240

Nearly complete backbone assignments were obtained using standard triple resonance experiments for both A. *a.* σ^{54} (60–135) and σ^{54} (69–198). The few missing resonances in σ^{54} (60–135) correspond to residues in a flexible linker between $\alpha 2$ and $\alpha 3$, residues 100 to 109. Interestingly, this linker could be assigned in the longer version (69–198) because the residues become more structured due to additional interactions, as will be discussed further below. In the 69–240 construct, many resonances could be assigned for residues 69–200, but residues C-terminal to 200 were very broad, missing, overlapped or otherwise unassignable. Only one segment of about 10 residues near the C-terminus could be assigned, and their chemical shift index values of these residues are consistent with a helix.

Structure of the σ^{54} (60–135) domain

The ensemble reflecting the solution structure of the σ^{54} (60–135) domain is shown in Figure 4A, and statistics for these 20 structures are given in Table 1. Residues 73–133 are well defined in the NMR solution structure with a root-mean-square deviation (RMSD) of 0.25 Å for the backbone atoms and 0.81 Å for heavy atoms. The N-terminal 13 residues are considerably less well-ordered, and the unassignable linker region between $\alpha 2$ and $\alpha 3$ is also less defined. σ^{54} (60–135) folds into a globular four-helix bundle, comprising amino acids 71–84 ($\alpha 1$), 90–99 ($\alpha 2$), 110–117 ($\alpha 3$), and 121–132 ($\alpha 4$). All of the helices are relatively short, with $\alpha 3$ having about two turns, and $\alpha 1$, $\alpha 2$, and $\alpha 4$ approximately three turns. $\alpha 1$ and $\alpha 2$ form an antiparallel helical hairpin, with a connecting five-residue loop. A longer loop (8 residues, less well-ordered than the secondary structure elements) connects to a short $\alpha 3$, followed by a short loop (3 residues) and $\alpha 4$. $\alpha 4$ packs across the closed end of the $\alpha 1$ – $\alpha 2$ hairpin, and $\alpha 3$ packs into the cleft formed by $\alpha 2$ and $\alpha 4$. The features of this fold are typical for helical bundle domains.

The packing brings together residues to form a well-defined hydrophobic core (comprised of 10 Leu, 2 Ile, 4 Val, 1 Phe, 1 Cys residues). The circular dichroism signal for this domain, typical of a helical fold, was followed as a function of temperature, but no melting transition was seen up to 80°C (data not shown). *A. aeolicus* is an extreme thermophile (growing optimally at 85°C), and thus we expect its proteins to be thermostable; however, it was surprising that this small domain alone is sufficient for this property.

This protein contains a total of 16 Asp/Glu residues and 8 Lys/Arg residues, giving a predicted net charge of –8 at neutral pH. The charged residues are spread throughout the protein's sequence and lead to a significant net negative electrostatic potential over almost the whole surface.

Structure of the extended σ^{54} (69–198) construct

Structure determination for σ^{54} (69–198) was performed in the same manner as for the shorter version (60–135). The ensemble representing the NMR solution structure of σ^{54} (69–198) is displayed in Figure 4B. Residues 73–186 have a backbone RMSD of 0.31 Å and an overall heavy atom RMSD of 0.81 Å, indicating that this fragment is very well-folded. Like σ^{54} (60–135), this protein has only helical secondary structure. The fold consists of seven helices, comprising residues 73–84 ($\alpha 1$), 88–99 ($\alpha 2$), 110–117 ($\alpha 3$), 121–133 ($\alpha 4$), 144–154 ($\alpha 5$), 160–172 ($\alpha 6$), and 178–186 ($\alpha 7$).

Within the σ^{54} (69–198) fragment, there are two distinct structural subdomains: the first, N-terminal subdomain includes $\alpha 1$ – $\alpha 4$ (the four helices that are in 60–135 remain in a similar packing arrangement), and the second C-terminal subdomain including $\alpha 5$ – $\alpha 7$, connected by a structured ten residue linker. The $\alpha 5$ – $\alpha 7$ subdomain helices form a three-helix bundle. $\alpha 5$ and $\alpha 6$ form an antiparallel helical hairpin, and $\alpha 7$ packs across them in a manner similar to $\alpha 1$, $\alpha 2$ and $\alpha 4$ (but with no analog to $\alpha 3$). The hydrophobic core of this three-helix bundle (comprised of 6 Leu, 4 Ile, 1 Phe, 1 Trp residues) is smaller than that for the $\alpha 1$ – $\alpha 4$ helices. The two subdomains have a substantial interface between them, primarily through hydrophobic interactions (with 3 Leu, 2 Val, 1 Phe, 1 Pro residues in this interface) and a potential salt bridge (R133 to E154, which are not highly conserved), Figure 5. The linker connecting the subdomains (residues 136–143) is structured and packed into the intersubdomain interface. The subdomain interface buries $\approx 975 \text{ \AA}^2$ of surface area, and this packing gives an overall structure that is elongated with a distinct bend.

Helix $\alpha 5$ and the N-terminal end of $\alpha 6$ are acidic, but the middle of $\alpha 6$ through the end of $\alpha 7$ is highly basic, with 12 total Lys/Arg residues and only 2 Glu/Asp residues. The basic residues are clustered together leading to a positively charged surface region on one end of the molecule, which could possibly interact with DNA in the promoter region.

Comparing the structures of the $\alpha 1$ – $\alpha 4$ region of σ^{54} 60–135 and 69–198 constructs

Comparison of the ^1H - ^{15}N HSQC spectra of σ^{54} (69–198) with the spectra of σ^{54} (60–135) revealed many chemical shift differences, despite the fact that σ^{54} (69–198) is more than 50% identical to σ^{54} (60–135). Indeed, overlaying the structures of σ^{54} (60–135) with the same residues within σ^{54} (69–198) gives a root mean square deviation for C_α atoms of 2.34 Å for helical regions. The overlay shows small changes in the relative orientations of helices between the two constructs, driven by interaction with the additional three helices in the longer construct. In the short construct, the loop between $\alpha 1$ and $\alpha 2$ is composed of five residues, while in the longer version there are just three connecting residues. Lys88 and Glu89 are part of the loop in σ^{54} (60–135) but part of helix $\alpha 2$ in σ^{54} (69–198).

The backbone could not be assigned for the linker between $\alpha 2$ and $\alpha 3$ in the short construct, indicating that the resonances are broadened by chemical exchange and that this segment is flexible with slow (ms- μ s) conformational dynamics. In σ^{54} (69–198), the broadening disappears, and this region exhibits a well-defined structure, stabilized by contacts between the $\alpha 2$ – $\alpha 3$ linker and the connector between the subdomains (residues 134–143). These interactions likely lead to the slight reorganization of the $\alpha 1$ to $\alpha 4$ helices in the longer construct.

Little structure is present for residues from 198 to 240

Secondary structure analysis of σ^{54} (69–240) based on both chemical shift indices and NOEs, showed that σ^{54} (69–240) contains the same folded residues as the shorter fragments. In addition, there is one extra region of helix toward the C-terminus that is connected to the entire seven-helical bundle by a disordered linker. Further, in the context of this longer fragment, there are modest changes in chemical shift within the folded 69–198 domain. These chemical shift differences indicate some interaction of this helix with the C-terminal part of the σ^{54} (69–198). However, no long-range contacts between this additional helix (residues 231–237) and other parts of the protein could be assigned, suggesting only a weak, transient interaction with this segment.

Interdomain interactions stabilize the folded state of residues 140–198

The spectra of the initial construct, σ^{54} (69–183), had considerably more variation in linewidth than the spectra of σ^{54} (69–198), and the number of observable peaks (>70%) was

considerably less than the expected 110. With the determined structure of σ^{54} (69–198), it is clear that the initial construct truncated $\alpha 7$, deleting two residues that pack into the core of the $\alpha 5$ – $\alpha 7$ subdomain and two residues that pack into the interface between subdomains. The spectra of σ^{54} (69–183) indicate that residues in $\alpha 5$ – $\alpha 6$ are relatively well-structured but that the destabilization from truncating $\alpha 7$ causes conformational exchange. This leads to line-width broadening and makes chemical shift assignments difficult.

To address whether the $\alpha 5$ – $\alpha 7$ region is in fact an independently folded domain, we studied σ^{54} (138–240) by NMR. The ^1H - ^{15}N HSQC spectrum of this construct, Figure 6A, has primarily very broad peaks, diagnostic of a poorly folded or unfolded structure. Thus, despite having a well-defined hydrophobic core, this segment of σ^{54} is not stable as an independent domain. However, when this ^{15}N -labeled protein was mixed with unlabeled σ^{54} (69–135) in a 1:1 ratio, the ^1H - ^{15}N HSQC spectrum changes dramatically, and many sharp peaks (≈ 50) appear, Figure 6B. This indicates an interaction between these two domains, even *in trans*, and partial stabilization of the folded structure of σ^{54} (138–240). Consistent with this, σ^{54} (69–135) and σ^{54} (138–240) coeluted from a size exclusion column (data not shown).

In the ^1H - ^{15}N HSQC spectrum of the complex, there is some variation in cross-peak intensity that may reflect residual effects of a chemical exchange process. The number of observed peaks is consistent with induced folding of the 138–198 region, corresponding to $\alpha 5$ – $\alpha 7$, but not residues further C-terminal in the protein. This is the same behavior seen with the 69–240 construct described above. Thus, the interdomain interaction is necessary to stabilize the $\alpha 5$ – $\alpha 7$ subdomain, and is sufficient to induce folding, even when the constructs are not covalently linked.

Discussion

Biochemical studies of σ^{70} identified segments of the protein important for particular functions, and limited proteolysis indicated the presence of several structural domains as well¹⁷. High-resolution x-ray crystal structures of σ^{70} in the context of the holopolymerase (*T.aquaticus*¹⁸; *T.thermophilis*²³) directly identified these domains, and the roles of the sequence elements were easily rationalized. The major domains for σ^{70} are (from N to C-terminus, residue numbers for the *T.t.* structure): ND1 (residues 74–254, interacting with core polymerase and DNA); ND2 (residues 258–313, also interacting with core polymerase, but occupying different positions in the *T.a.* and *T.t.* structure); LD (residues 314–340, a linker domain with little secondary structure); and CD (residues 341–423, which is responsible for sequence specific DNA binding at the –35 region of the promoter). Residues 1–73 play a role in inhibition of promoter binding, and a recent NMR solution structure for this region (residues 29–95 of *T. maritima* σ^A) shows that it also folds into a helical bundle like the other σ^{70} domains²⁴. Both structural and biochemical data support the idea that these functional domains rearrange during the processes of promoter recognition and opening⁵. Although there is no recognizable sequence homology between σ^{54} and σ^{70} , σ^{54} carries out functions analogous to those of σ^{70} , and biochemical data indicate that it too has functional domains^{10; 22; 25; 26}. The functional regions within σ^{54} are indicated schematically in Figure 1, and were described in the Introduction. The long C-terminal segment, known as Region III, contains core polymerase binding and DNA binding activities, and is comprised of multiple structural domains. The N-terminal part of Region III is the focus of this work. This segment of σ^{54} has been shown to be largely responsible for interaction with the core polymerase, but may also affect other interactions within the polymerase-DNA complex. The C-terminal end of σ^{54} is functionally analogous to the CD domain of σ^{70} in providing sequence specific DNA binding, and was shown to be structurally similar as well¹³.

Boundaries for the folded core polymerase binding domain of σ^{54}

In this work we produced constructs of σ^{54} containing sequences that have been shown by prior deletion analysis and mutagenesis studies to interact with the core RNA polymerase^{12; 21; 27; 28; 29}. The most extensive deletion mapping²⁹ indicated that the key elements for core binding are between residues 74–158 (corresponding to 120–215 in *Klebsiella pneumoniae* which Gallegos and Buck studied). This region corresponds with residues comprising $\alpha 1$ through $\alpha 5$ defined by our data, including the stable four-helix bundle core.

The Pfam database identifies a longer segment, residues 60–235, as the “core binding domain”, top row in Figure 2. This domain assignment seems to have been based on biochemical data only⁴². The sequence in this region of the protein is not highly conserved, although there are some completely conserved residues and regions of significant sequence similarity. The structural data, discussed in further detail below, show that residues 69–135 alone correspond to a well-folded, stable four-helix bundle. Extension to include residues 136–198 adds three additional helices that fold when packed against residues 69–135. The well-ordered part of the longer constructs includes residues 69–193 (first and last well-ordered residues). Inspection of the secondary structure, shown above the sequence in Figure 2, indicates that the C-terminal end of the domain that was defined through genetically introduced deletions occurs just before helix-6. The conservative substitutions for many hydrophobic residues in this region (discussed further below) indicate that the folded cores of the *Aquifex* and *Klebsiella* proteins are probably the same.

Structure of the N-terminal 4 helix bundle, $\sigma^{54}(60-135)$

By comparing different length constructs of σ^{54} Region III, we identified residues 60–135 as forming a well-folded domain. The NMR structure presented in the Results section shows that this domain has four helices, packed into a tight, basically spherical bundle. A Dali search returned nearly 400 structures with a Z-score > 2. One of the top two hits (Z=5.0, 2.7 Å rmsd) was pdb 1KW4, a polyhomeotic SAM domain structure, a type of domain frequently involved in protein-protein interactions in eukaryotes and often part of a larger multi-domain protein³⁰. It has the same basic topology, with two small additional helical segments relative to $\sigma^{54}(69-135)$. Since there are so many proteins with approximately this fold, a structure comparison with them does not provide much information about a specific function.

As described in the Results section, there is a cluster of residues that forms a well-defined hydrophobic core for this structure, stabilizing a fold with a melting transition above 80°C. A few of these residues are absolutely conserved in σ^{54} , and the remaining are almost always hydrophobic, Figure 2. This domain is highly charged, with a total of 24 acidic and basic residues (15 Glu, 1 Asp, 6 Lys and 2 Arg), giving a net –8 charge at neutral pH. These residues are dispersed over the surface. The presence of charged residues has been associated with thermal stability in some cases³¹; however, the mesophile versions are almost as highly charged, and the overall distribution of charges is similar in thermo- and mesophilic versions of the protein. Hence it is unclear how much charged residues contribute to stability in this case.

Structure of $\sigma^{54}(69-198)$

The C-terminally elongated construct (69–198) retains the basic fold of residues 69–135 with the addition of three more helices. These C-terminal helices pack together in a fold somewhat like that of 69–135, forming a well-defined hydrophobic core with residues that are not absolutely conserved but are hydrophobic in homologs, Figure 1. The extension, residues 136–198, packs against the N-terminal four-helix bundle. $\alpha 5$ contacts the end of $\alpha 4$,

but the interface is created primarily by the residues between $\alpha 4$ and $\alpha 5$. These are well-ordered, making a helical turn. There is significant sequence conservation of these interfacial residues, particularly PxG in which P and G are completely conserved. This interface also contains some residues immediately C-terminal of $\alpha 7$ (including highly conserved P191) and residues in the $\alpha 2$ to $\alpha 3$ linker, which changes conformation significantly between the 60–135 and 69–198 structures. Small changes in the relative orientations of the helices within the four-helix bundle, particularly $\alpha 3$, also occur upon packing with the linker and additional helices.

The combination of the two subdomains packed together gives rise to a distinct bend in the structure, which is apparent in Figures 4 and S3. A Dali search identified only folds homologous to the subdomains, not to the overall organization of the full structure. This may be a reflection of the unusual packing between the four-helix bundle and the three-helix subdomain.

Further extension of the expressed sequence to residue 240 did not yield additional well-folded regions, although there seems to be one additional helical element near the end of the longer construct, from Glu231 to Leu237. Modest changes in chemical shifts in 69–198 relative to 69–240 are observed, but no NOE contacts to 231–237 were detected, suggesting only weak interactions with this extension.

In the context of the 69–198 construct, the $\alpha 5$ - $\alpha 6$ - $\alpha 7$ bundle is as well-defined as the N-terminal four-helix bundle, and the two are well-ordered with respect to one another. However, when residues 138–240 (beginning of $\alpha 5$ through the end of the domain) were expressed as an independent construct, they did not show a stable, folded structure. Peaks in the NMR spectra of 138–240 (Figure 6, left) were generally very broad, indicating slow conformational fluctuations, perhaps between unfolded and partially folded states. We cannot exclude that this construct aggregates, causing broadening of resonances, but the appearance of some peaks with a range of linewidths makes this seem unlikely. When the 138–240 polypeptide was mixed *in trans* with the N-terminal domain, residues 60–135, the three-helical subdomain did fold, yielding sharp, well-dispersed resonances (Figure 6, right). This suggests that the combination of the hydrophobic core of the three-helix bundle and the interface with the four-helix bundle (which is $\approx 975 \text{ \AA}^2$) is sufficient to stabilize this subdomain, while the hydrophobic core alone is not. This is an unusual feature of the structure. The conservation of hydrophobic residues involved in stabilization of both the three-helix bundle and the interface between the subdomains (Figure 2) indicates that this basic architecture is conserved in σ^{54} and suggests that it may have functional importance.

Site of the CBD in the polymerase complex

Recently Bose *et al.* published a low resolution, electron microscopy-based structure of the holopolymerase with σ^{54} alone and in complex with the activator PspF³². In this structure, there are three regions of density that protrude from the holopolymerase, which are referred to as D1–D3, and a ‘bridging’ density, Db. D3 was clearly identified as the DNA binding domain of σ^{54} (the C-terminal end of Region III), and Db is thought to be Region I. Given these assignments, the density in D1 is most likely the core binding domain, corresponding to the folded region 69–198 of σ^{54} presented here. The size of the density in this region is approximately correct for this domain. With further information about domain structures, such as that reported here for the core binding domain, it may be possible to model the sigma factor in the complex. The D1 density changes location somewhat in the complex with PspF. In their model, which included DNA, the D1 density is positioned near the start site for transcription, suggesting that the positively charged region of $\alpha 6$ and $\alpha 7$ could be in a position to make non-specific DNA contacts.

Interpretation of mutagenesis data

Prior biochemical studies found that residues 73–158²⁹ are sufficient to bind core polymerase tightly, and therefore, this region is called the “core binding domain.” From the sequence alignment, this corresponds to helices $\alpha 1$ through $\alpha 5$ in our determined structure. This region has all residues required to form the N-terminal four-helix bundle but is missing most of the residues that form the three-helix segment that packs against it. This implies that the four-helix bundle contains most of the contacts to the core polymerase. However, Gallegos and Buck made closely related constructs that produce results not predicted by our structure. They found that a slightly shorter construct, corresponding to residues 73–151 (residues 120–208 in *K.p.*), lost significant binding affinity for core. The seven residues that were deleted in that case correspond to the linker between $\alpha 5$ and $\alpha 6$ in our structure. These residues are not highly conserved and in fact five of these residues have been evolutionarily deleted in the *A.a.* protein according to the alignment, Figure 2, and hence it is difficult to rationalize why removal of residues 152–158 would affect folding or function. Other constructs made by Gallegos and Buck show some coupling in effects when both N- and C-terminal changes to the folded four-helix bundle are made, which again cannot be explained by our structures. Almost all of their deletion constructs had N-terminal His₆ tags; it is possible that this could interact with the negative surface of the four-helix bundle and affect its activity.

Mutagenesis studies also identified the same region of σ^{54} as important for core polymerase binding. Gralla and coworkers²⁷ noted a region of the sequence with hydrophobic and acidic repeats (residues 109–128). Internal sequence deletions in this region (removing 122–139 or 127–131) were defective in core polymerase binding, which is consistent with our observations since the deletions removed all or part of $\alpha 4$ in the four-helix bundle and would be expected to completely disrupt the structure. The hydrophobic residues, which were mutated to Ser, are important in forming the hydrophobic core of the four-helix bundle, and the loss of activity again probably arose from disruption of the fold. Mutations of the acidic residues were performed in groups of three (Glu to Gln and Asp to Asn), and had less dramatic effects. The altered residues are all exposed, and the changes should not affect folding. Changing six acidic residues at once does reduce activity, which may be a result of lost negative electrostatic potential that could contribute to binding.

These workers also noted a sequence similarity to a part of σ^{70} , and although it appeared not to be statistically significant at the sequence level, they thought there might be some analogy. In σ^{70} this sequence is a long helix with a slight kink; whereas in σ^{54} (residues 112 to 138) it corresponds to $\alpha 3$, all of $\alpha 4$, and a few residues of the linker to $\alpha 5$. Though this region is important for folding in each case, the display of sidechains is very different between the two, and it seems unlikely that there is any functional similarity in the interactions of these residues with other parts of polymerase.

In addition to point mutations directed to the hydrophobic/acidic repeats, Tintut and Gralla also carried out extensive random mutagenesis to identify functionally significant residues²¹. Residues that cannot be changed without loss of function map primarily to the four-helix bundle region in our structure (between residues 106 and 138). Among the twelve sites they identified in two screens seven are hydrophobic (Leu106, Leu117, Cys119, Leu124, Val127, Val131, Leu132) and well-buried in the structure. Another one is a glycine deeply buried in the interface between the four-helix and three-helix bundles (Gly138). Changes in these presumably disrupt the fold, and hence also the function. Two more of the residues were hydrophobes in the mutagenized *E. coli* protein (Ala177 and Leu179), but are charged (Lys126 and Arg128) in *A. aeolicus*. The position of Lys126 is surface exposed. The backbone of the Arg128 residue is fairly buried in our structures, but the sidechain is not well defined by the data and may be exposed at the surface. It is difficult to interpret

these observations. The final two residues for which changes caused loss of function are surface exposed, charged residues: Asp115 (Glu163 in *E.c.*) and Lys130 (Arg181 in *E.c.*). The charge (but not absolute identity) is retained in most σ^{54} sequences, and makes them likely candidates for interaction with the core polymerase. They are on somewhat different parts of the four-helix bundle, but both could easily be in contact with the core although presumably they could interact with other parts of σ^{54} as well.

Comparison with the core binding region of σ^{70}

Region ND1 of σ^{70} provides a major surface of contact with the core polymerase enzyme⁵. ND1 is somewhat larger than the core binding domain of σ^{54} (ca. 180 residues rather than 130), and the folding topology of the two proteins is completely different. However, like the core binding domain of σ^{54} , ND1 is helical with two lobes, Figure 7A.

The CBD of σ^{54} has a somewhat similar shape as the base of σ^{70} ND1; it is negatively charged like ND1; and, it was identified as important for core binding by mutagenesis. Currently available data are not sufficient to define the position of the σ^{54} CBD with respect to the core, but it would fit into a region similar to the base of ND1.

Role for a possible ‘fracture point’ in σ^{54}

The structure of the σ^{54} CBD is unusual in having a stable four-helix bundle with a second subdomain that is not intrinsically stable but that folds through interaction with it. This behavior is consistent with the moderate size hydrophobic interface between these structure elements. The relatively high sequence conservation in the region of this interdomain interface strongly suggests that it is a conserved structural feature of σ^{54} 's core binding domain.

The σ^{54} CBD connects Region I on its N-terminal end, which is responsible for activator interactions, to the domain that contacts DNA in the –12 region of the promoter where DNA opening initiates. When catalytically active, the ATPase domains of the activators form rings with loops extending above the ring. These loops make contact with Region I of σ^{54} , and upon ATP hydrolysis, the loops are pulled flat, presumably pulling on Region I of σ^{54} and then releasing it³³. This step drives a conformational change in the σ^{54} -core polymerase-DNA complex that enables the opening of DNA and transcription initiation. The interface between the four-helix bundle and three-helix extension in the σ^{54} CBD domain could play the role of a designed fracture point, a specific weak point in the structure intended to anchor the domains together until a stress is applied from the ATPase domain. The resulting conformational change of the CBD could then expose new interaction surfaces, causing the reorganization of polymerase to the state competent for opening DNA and initiating transcription. The structure of the CBD reported here will facilitate design of experiments to test for such a role.

Materials and Methods

Protein purification

DNA encoding fragments of *Aquifex aeolicus* σ^{54} (specifically residues 60–135, 69–135, 69–198, 69–240 and 138–240) were cloned into pET21b3-2 (Pharmacia Biotech) with a (His)₆-tag at the N-terminus followed by a TEV cleavage site. The fragments were expressed in *E. coli* BL21(DE3) with the Rosetta. pLysS plasmid. Cells were grown on standard minimal medium with ¹⁵N NH₄Cl and/or ¹³C glucose. Protein expression was induced at A_{600 nm} = 0.5–0.6 with 1 mM isopropyl β-D-thiogalactopyranoside. Five hours later, cells were harvested by centrifugation and suspended in 20 mM sodium phosphate (pH 7.4), 0.5 M NaCl, and 2 mM β-mercaptoethanol. After sonication, the cell lysates were

centrifuged at 30,000g for 30 min, and the supernatants filtered (0.45 μ m). The (His)₆tag-proteins were loaded onto a HiTrap^{TM} Chelation HP column (Pharmacia Biotech) and eluted with a continuous 100 mL linear gradient of imidazole from 0 to 500 mM. His-tags were cleaved by incubation with TEV protease for 12 hr in batch conditions. Proteins were further purified by gel filtration column chromatography on a HiLoad Superdex 75 preparative column (Pharmacia Biotech) with a running buffer of 25 mM sodium phosphate (pH 7.0), 100 mM NaCl, and 2 mM dithiothreitol (pH 7.0).

Refolding from inclusion bodies

For the σ^{54} (138–240) construct, protein expression generated inclusion bodies. After sonication and centrifugation, the inclusion bodies were solubilized by stirring for 12 hr at 4°C in 8 M urea containing Tris-HCl (pH 7.5), 0.5 M NaCl and 2 mM β -mercaptoethanol. The solution was then loaded onto a HiTrap^{TM} Chelation HP column (Pharmacia Biotech) and refolded using a linear urea gradient (0.2 mL/min) from 8 M to 0 M over 100 mL. The protein was eluted with a continuous 100 mL linear gradient of imidazole from 0 to 500 mM and then dialyzed overnight against 25 mM sodium phosphate (pH 7.0), 100 mM NaCl, and 2 mM dithiothreitol (pH 7.0). The protein was then passed through a HiLoad Superdex 75 preparative column (Pharmacia Biotech).

NMR spectroscopy

All NMR experiments were recorded at 303 K on Bruker DRX800 and 900 MHz spectrometers equipped with z-shielded gradient triple resonance cryoprobes. ¹H chemical shifts were referenced directly to the methyl proton signals of DSS at 0 ppm, and ¹³C and ¹⁵N chemical shifts calculated indirectly from the ¹H spectrometer frequency and known ratios of the reference γ values. Spectra were processed with the NMRPipe/nmrDraw package³⁴ and analyzed using the program Sparky³⁵. Three-dimensional experiments for sequential assignments³⁶ included HNCACB, CBCA(CO)NH, C(CCO)NH, H(CCO)NH, HBHA(CBCACO)NH and HCCH-TOCSY experiments. For restraint generation, ¹⁵N-edited HSQC-NOESY (τ_m = 120ms) and ¹³C-edited NOESY-HSQC (τ_m = 100ms) spectra were acquired. Dihedral backbone constraints were derived from ¹H _{α} , ¹³C _{α} , ¹³C _{β} chemical shifts with the program TALOS³⁷. Hydrogen bonds were identified based on slowly exchanging amide protons observed by recording ¹H-¹⁵N HSQC spectra immediately after dissolving lyophilized ¹⁵N-labeled protein in D₂O.

Structure calculations

Structure calculations on σ^{54} domains were performed using a semi-automated process in CYANA 2.1³⁸. Unambiguous sidechain NOE cross-peak assignments were obtained manually and used in an initial round of structure calculations to identify the secondary structure topology. These initial structures were then used to filter possible assignments for previously ambiguous peaks. Seven cycles of CYANA generated further sidechain assignments, each cycle having 10000 steps of torsion-angle dynamics with a simulated annealing protocol. A set of 100 structures was calculated, and the 20 structures with the lowest target function values were chosen for analysis. The final structures were validated with PROCHECK³⁹. Structures were analyzed and visualized using PyMOL⁴⁰ and MOLMOL⁴¹. The electrostatic surface potential was calculated with APBS⁴². The 20 analyzed structures were deposited into the PDB for each construct.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH grant GM 62163 to D.E.W., and partially by a fellowship from the Korea Research Foundation (MOEHRD) KRF-2006-214-C00056, to E.H. The authors thank J. Pelton for help with NMR spectroscopy, and S. Kustu and B.T. Nixon for helpful discussions. NMR spectrometers used for this work were supported by NSF and NIH (grants NSF BBS 01-19304, NIH RR15756 and NIH-GM68933).

References

1. Cramer P. Multisubunit RNA polymerases. *Curr Opin Struct Biol.* 2002; 12:89–97. [PubMed: 11839495]
2. Kornberg RD. The molecular basis of eukaryotic transcription. *Proc Natl Acad Sci U S A.* 2007; 104:12955–61. [PubMed: 17670940]
3. Burgess RR, Travers AA, Dunn JJ, Bautz EK. Factor stimulating transcription by RNA polymerase. *Nature.* 1969; 221:43–6. [PubMed: 4882047]
4. Hirschman J, Wong PK, Sei K, Keener J, Kustu S. Products of nitrogen regulatory genes *ntrA* and *ntrC* of enteric bacteria activate *glnA* transcription in vitro: evidence that the *ntrA* product is a sigma factor. *Proc Natl Acad Sci U S A.* 1985; 82:7525–9. [PubMed: 2999766]
5. Murakami KS, Darst SA. Bacterial RNA polymerases: the whole story. *Curr Opin Struct Biol.* 2003; 13:31–9. [PubMed: 12581657]
6. Mooney RA, Darst SA, Landick R. Sigma and RNA polymerase: an on-again, off-again relationship? *Mol Cell.* 2005; 20:335–45. [PubMed: 16285916]
7. Gallegos MT, Buck M. Sequences in sigma(54) region I required for binding to early melted DNA and their involvement in sigma-DNA isomerisation. *J Mol Biol.* 2000; 297:849–59. [PubMed: 10736222]
8. Fisher MA, Grimm D, Henion AK, Elias AF, Stewart PE, Rosa PA, Gherardini FC. Borrelia burgdorferi sigma54 is required for mammalian infection and vector transmission but not for tick colonization. *Proc Natl Acad Sci U S A.* 2005; 102:5162–7. [PubMed: 15743918]
9. Saldias MS, Lamothe J, Wu R, Valvano MA. Burkholderia cenocepacia requires the RpoN sigma factor for biofilm formation and intracellular trafficking within macrophages. *Infect Immun.* 2008; 76:1059–67. [PubMed: 18195023]
10. Cannon WV, Chaney MK, Wang X, Buck M. Two domains within sigmaN (sigma54) cooperate for DNA binding. *Proc Natl Acad Sci U S A.* 1997; 94:5006–11. [PubMed: 9144180]
11. Casaz P, Gallegos MT, Buck M. Systematic Analysis of s⁵⁴ N-terminal Sequences Identifies Regions Involved in Positive and Negative Regulation. *J Mol Biol.* 1999; 292:229–239. [PubMed: 10493871]
12. Wong C, Tintut Y, Gralla JD. The domain structure of sigma 54 as determined by analysis of a set of deletion mutants. *J Mol Biol.* 1994; 236:81–90. [PubMed: 8107127]
13. Doucleff M, Pelton JG, Lee PS, Nixon BT, Wemmer DE. Structural basis of DNA recognition by the alternative sigma-factor, sigma54. *J Mol Biol.* 2007; 369:1070–8. [PubMed: 17481658]
14. Wigneshweraraj SR, Burrows PC, Bordes P, Schumacher J, Rappas M, Finn RD, Cannon WV, Zhang X, Buck M. The second paradigm for activation of transcription. *Prog Nucleic Acid Res Mol Biol.* 2005; 79:339–69. [PubMed: 16096032]
15. Jain D, Nickels BE, Sun L, Hochschild A, Darst SA. Structure of a ternary transcription activation complex. *Mol Cell.* 2004; 13:45–53. [PubMed: 14731393]
16. Sorenson MK, Ray SS, Darst SA. Crystal structure of the flagellar sigma/anti-sigma complex sigma(28)/FlgM reveals an intact sigma factor in an inactive conformation. *Mol Cell.* 2004; 14:127–38. [PubMed: 15068809]
17. Campbell EA, Muzzin O, Chlenov M, Sun JL, Olson CA, Weinman O, Trester-Zedlitz ML, Darst SA. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell.* 2002; 9:527–39. [PubMed: 11931761]
18. Murakami KS, Masuda S, Campbell EA, Muzzin O, Darst SA. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science.* 2002; 296:1285–90. [PubMed: 12016307]

19. Svergun DI, Malfois M, Koch MH, Wigneshweraraj SR, Buck M. Low resolution structure of the sigma54 transcription factor revealed by X-ray solution scattering. *J Biol Chem.* 2000; 275:4210–14. [PubMed: 10660585]
20. Rappas M, Schumacher J, Beuron F, Niwa H, Bordes P, Wigneshweraraj S, Keetch CA, Robinson CV, Buck M, Zhang X. Structural insights into the activity of enhancer-binding proteins. *Science.* 2005; 307:1972–75. [PubMed: 15790859]
21. Tintut Y, Gralla JD. PCR mutagenesis identifies a polymerase-binding sequence of sigma 54 that includes a sigma 70 homology region. *J Bacteriol.* 1995; 177:5818–25. [PubMed: 7592329]
22. Paget MSB, Helmann JD. the σ^{70} family of sigma factors. *Genome Biol.* 2003; 4:203. [PubMed: 12540296]
23. Vassilyev DG, Sekine S, Laptenko O, Lee J, Vassilyeva MN, Borukhov S, Yokoyama S. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature.* 2002; 417:712–9. [PubMed: 12000971]
24. Schwartz EC, Shekhtman A, Dutta K, Pratt MR, Cowburn D, Darst S, Muir TW. A full-length group 1 bacterial sigma factor adopts a compact structure incompatible with DNA binding. *Chem Biol.* 2008; 15:1091–103. [PubMed: 18940669]
25. Sasse-Dwight S, Gralla JD. Role of eukaryotic-type functional domains found in the prokaryotic enhancer receptor factor sigma 54. *Cell.* 1990; 62:945–54. [PubMed: 2203540]
26. Missailidis S, Cannon WV, Drake A, Wang XY, Buck M. Analysis of the architecture of the transcription factor sigma N (sigma 54) and its domains by circular dichroism. *Mol Microbiol.* 1997; 24:653–64. [PubMed: 9179857]
27. Tintut Y, Wong C, Jiang Y, Hsieh M, Gralla JD. RNA polymerase binding using a strongly acidic hydrophobic-repeat region of sigma 54. *Proc Natl Acad Sci U S A.* 1994; 91:2120–4. [PubMed: 8134358]
28. Cannon W, Missailidis S, Smith C, Cottier A, Austin S, Moore M, Buck M. Core RNA polymerase and promoter DNA interactions of purified domains of sigma N: bipartite functions. *J Mol Biol.* 1995; 248:781–803. [PubMed: 7752240]
29. Gallegos MT, Buck M. Sequences in sigmaN determining holoenzyme formation and properties. *J Mol Biol.* 1999; 288:539–53. [PubMed: 10329161]
30. Qiao F, Bowie JU. The many faces of SAM. *Sci STKE.* 2005; 2005:re7. [PubMed: 15928333]
31. Haney P, Konisky J, Koretke KK, Luthey-Schulten Z, Wolynes PG. Structural basis for thermostability and identification of potential active site residues for adenylate kinases from the archaeal genus *Methanococcus*. *Proteins.* 1997; 28:117–30. [PubMed: 9144797]
32. Bose D, Pape T, Burrows PC, Rappas M, Wigneshweraraj SR, Buck M, Zhang X. Organization of an activator-bound RNA polymerase holoenzyme. *Mol Cell.* 2008; 32:337–46. [PubMed: 18995832]
33. Chen B, Doucleff M, Wemmer DE, De Carlo S, Huang HH, Nogales E, Hoover TR, Kondrashkina E, Guo L, Nixon BT. ATP ground- and transition states of bacterial enhancer binding AAA+ ATPases support complex formation with their target protein, sigma54. *Structure.* 2007; 15:429–40. [PubMed: 17437715]
34. 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; 6:277–93. [PubMed: 8520220]
35. Goddard, T.; Kneller, DG. SPARKY. University of California; San Francisco: 2004.
36. Ferentz AE, Wagner G. NMR spectroscopy: a multifaceted approach to macromolecular structure. *Quart Rev Biophys.* 2000; 33:29–65.
37. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR.* 1999; 13:289–302. [PubMed: 10212987]
38. Guntert P. Automated NMR structure calculation with CYANA. *Methods Mol Biol.* 2004; 278:353–78. [PubMed: 15318003]
39. 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; 8:477–86. [PubMed: 9008363]

40. DeLano WL. Use of PYMOL as a communications tool for molecular science. Abstr Pap Am Chem S. 2004; 229:U313–U314.
41. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph. 1996; 14:51–5. 29–32. [PubMed: 8744573]
42. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A. 2001; 98:10037–41. [PubMed: 11517324]
43. Buck M, Gallegos M-T, Studholme DJ, Guo J, Gralla D. The Bacterial Enhancer-Dependent σ^{54} (σ^N) Transcription Factor. J Bact. 2000; 182:4129–4136. [PubMed: 10894718]

Figure 1.

The sequence of σ^{54} is indicated by the broad line. Various regions are indicated by color: yellow, interaction with enhancer binding proteins; blue, core binding domain; orange, interaction with DNA in the -12 region of the promoter (for which precise domain boundaries have not been identified); red, sequence specific recognition of the -24 promoter region.

Figure 2.

An alignment of σ^{54} protein sequences for the region defined by Pfam as the core binding domain from *A. aeolicus* (top) and ten other diverse organisms. Conserved residues in structured regions are highlighted: turquoise indicates charged residues; green indicates hydrophobic residues (in *A.a.* the residues involved in formation of the hydrophobic cores); dark yellow indicates residues in the interface between the four-helix bundle and the three-helix subdomain. Completely conserved residues are in bold lettering with an * at the bottom.

Figure 3. ^1H - ^{15}N HMQC spectra for the 60–135 construct (top) and the 69–198 construct (bottom), with peaks labeled by the residue assignments. Sidechain amide resonances of Asn and Gln are indicated with horizontal lines.

Figure 4.

Ensembles of the 20 best solution structures (stereo) determined from NMR data for (A) construct $\sigma^{54}(60-135)$ (disordered residues 60–67 removed for clarity) and (B) construct $\sigma^{54}(69-198)$. The orientation of the four-helix bundle in (B) is the same as the domain in (A). Helices are colored blue and are numbered as referred to in the text.

Figure 5.

Residues in the interface between the four-helix bundle (lower part of the structure) and three-helix extension (upper part of the structure) are shown as sticks. Atoms in these contact residues are colored according to atom type, other residues are blue (helix) or green (coil) depending on secondary structure; a possible salt bridge is shown with the sidechains in atom colors (left center), Glu 154 behind Arg133.

Figure 6. ^1H - ^{15}N HSQC spectra are shown for ^{15}N -labeled σ^{54} residues 138–240 in the absence of any other polypeptide (left) and in the presence of unlabeled σ^{54} (60–135) (right).

Figure 7. A comparison of the core polymerase binding regions of σ^{54} CBD (left) and σ^{70} ND1 (right) as individual domains. Although these have different folding topologies, there is a similarity in shape. Non-core residues that were identified as important for function of the σ^{54} CBD are highlighted in orange.

Table 1 Structural statistics for the final calculated structures of the s⁵⁴ 69–135 and s⁵⁴ 69–198 proteins

Restraints	60–135	69–198
All	1572	3224
Intraresidue	408	926
Sequential ($ i-j = 1$)	412	821
Medium range ($2 \leq i-j \leq 5$)	466	861
Long range ($ i-j > 5$)	286	616
Hydrogen bond distance restraints ^a	29	44
Backbone dihedral angles		
All	67	190
Phi	33	93
Psi	34	97
Violations		
Distances $> 0.5 \text{ \AA}$	None	None
Dihedral angles $> 5^\circ$	None	4
Average target function ($\text{\AA}^2$)	1.45 ± 0.11	7.87 ± 0.34
Mean RMS deviations from the average^b		
Backbone atoms (N, C _{α} , C', O)	$0.25 \pm 0.05 \text{ \AA}$	$0.31 \pm 0.10 \text{ \AA}$
Heavy atoms	$0.81 \pm 0.06 \text{ \AA}$	$0.81 \pm 0.06 \text{ \AA}$
Ramachandran statistics		
Most favored regions	82.2%	85.9%
Additional allowed regions	17.6%	12.6%
Generously allowed regions	0.1%	1.5%
Disallowed regions	0%	0%

^aTwo restraints per hydrogen bond.^bRmsd values for residues 73–133 for s⁵⁴ 60–135 and 73–186 for s⁵⁴ 69–198.