THE C-TERMINAL RPON DOMAIN OF σ^{54} FORMS AN UNPREDICTED HELIX-TURN-HELIX MOTIF SIMILAR TO DOMAINS OF σ^{70}

Michaeleen Doucleff[1,2], Lawrence T. Malak[1,3], Jeffrey G. Pelton[1,2], David E. Wemmer*[1,2]

From [1]Physical Biosciences Division, Lawrence Berekeley National Laboratory, [2]Department of Chemistry, and [3]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA
* Correspondence should be directed to David E. Wemmer: Physical Biosciences Divisions, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA; e-mail: DEWemmer@lbl.gov; fax: (510) 486-6059.

The 'σ' subunit of prokaryotic RNA-polymerase allows gene-specific transcription initiation. Two σ families have been identified, σ^{70} and σ^{54}, which use distinct mechanisms to initiate transcription and share no detectable sequence homology. Although the σ^{70}-type factors have been well characterized structurally by x-ray crystallography, no high-resolution structural information is available for the σ^{54}-type factors. Here we present the NMR derived structure of the C-terminal domain of σ^{54} from *Aquifex aeolicus*. This domain (Thr323 to Gly389), which contains the highly conserved RpoN box sequence, consists of a poorly structured N-terminal tail followed by a three-helix bundle, which is surprisingly similar to domains of the σ^{70}-type proteins. Residues of the RpoN box, which have previously been shown to be critical for DNA binding, form the second helix of an unpredicted helix-turn-helix motif. This structure's homology with other DNA binding proteins, combined with previous biochemical data, suggest how the C-terminal domain of σ^{54} binds to DNA.

INTRODUCTION

Transcription, the synthesis of RNA from double-stranded DNA, is a fundamental process in all forms of life. The primary protein complex catalyzing transcription is composed of five subunits, $\alpha_2\beta'\beta\omega$, called 'core RNA polymerase (RNAP)'. Core RNAP is fully competent to synthesize RNA from DNA. However, the initiation of transcription at specific DNA sequences requires additional protein(s). In bacteria, these additional proteins are called the 'σ factors' (1, 2). The roles of the σ factor are three fold: 1. Bind to the core RNAP to form the σ-RNAP holoenzyme; 2. Recognize and bind to a specific DNA sequence next to the transcription start site, called the promoter element; 3. Facilitate the initiation of transcription by opening the double-stranded DNA.

Based on protein sequence homology, σ factors can be grouped into two classes, σ^{70} and σ^{54}, which were named by the molecular weights of the first members identified (reviewed in 3). The σ^{70}-type factors are the most abundant in bacteria (reviewed in 4, 5). They include the primary σ factors, such as σ^{70} from *Escherichia coli* and σ^{A} from Gram-positive bacteria, which regulate transcription for most genes required for normal exponential growth. Other members of the σ^{70} family (σ^{28}, σ^{32}, et cetera) regulate the transcription of more specialized genes that are required to respond to environmental changes. For example, σ28 from *Salmonella typhimurium* controls expression of genes required for flagellar assembly (6).

The σ^{54}-type factor, which is also called σ^{N} and encoded by the *rpoN* gene, has no detectable sequence homology to σ^{70}-type factors. Its occurrence is widespread among bacteria, but there is usually only one σ^{54}gene present in a particular organism (7). σ^{54} regulates gene transcription for proteins required for many important cellular functions, such as nitrogen metabolism, chemotaxis, and heat and phage shock responses. Recently, σ^{54} has been shown to be required for mammalian infection by *Borrelia burgdorferi*, the agent of Lyme disease (8).

The most striking difference between σ^{54} and σ^{70} factors is how they regulate transcription

initiation. Once σ^{70}-RNAP holoenzyme binds to the promoter of a gene (at a conserved sequence -10 and -35 base pairs upstream of the transcription start site), it can open DNA and transcription can begin spontaneously. In contrast, when σ^{54}-RNAP holoenzyme binds to a promoter (at a consensus sequence -12 and -24 base pairs upstream of the transcription start site), it remains in the closed, inactive state. To open the double stranded DNA and intiate transcription, σ^{54}-RNAP holoenzyme requires interaction with an activator protein, (reviewed in 9), which binds approximately 150 base pairs upstream of the promoter (10). Once stimulated by a physiological signal, such as a phosphorylation or ligand binding, the activator uses the energy of ATP hydrolysis (via a conserved AAA+-ATPase domain) to remodel σ^{54}-RNAP into an active conformation that is capable of initiating transcription (7).

The $\sigma\sigma^{70}$-type factors have been well characterized structurally (11, 12, 13, 14). X-ray crystal structures of individual domains σ^{70} and σ^{70}-RNAP holoenzyme show that σ^{70}-type factors are composed of four helical domains (σ^1, σ^2, σ^3, σ^4) connected by flexible linkers (15, 16). These four domains are also functionally distinct: σ^1 is involved in regulating the kinetics of transcription initiation; σ^2 binds to the -10 promoter element and is essential for melting the DNA; σ^3 stabilizes the open complex formation by binding the extended -10 element; and, σ4 binds specifically to the -35 promoter element (reviewed in 17).

In contrast to σ^{70}, σ^{54} is divided into three regions based on function (18, 19, 20): region 1 (*E. coli* 1-55) interacts with the upstream activator protein to control promoter melting; region 2 (*E. coli* 70-304) contains the minimal RNAP binding domain (*E. coli* 70-180) and a part which enhances DNA binding affinity (*E. coli* 180-304); Region 3 (*E. coli* 329-477) recognizes and binds the consensus promoter elements. This C-terminal part of the protein contains a region which crosslinks to DNA (*E. coli* 329-346), a predicted helix-turn-helix motif (*E. coli* 366-386; Figure 1), and a highly conserved sequence (*E. coli* 454-463; Figure 1) called the RpoN box. The RpoN box (ARRTVAKYRE) is the signature sequence for σ^{54} proteins and is critical for DNA binding (21).

Although information about the overall shape of σ^{54} has come from scattering and electron micscropy data (22, 23), no high-resolution structural information is available for σ^{54}. The lack of sequence homology between σ^{54} and σ^{70} prevents homology modeling of σ^{54} domains based on the structures of σ^{70} domains. Therefore, to understand the mechanism of σ^{54}-dependent transcription initiation, high-resolution structures of domains are required.

Here we present the NMR derived structure of a C-terminal fragment of σ^{54} (residues Thr323 to Gly389) from the hyperthermophilic bacterium, *Aquifex aeolicus*. This domain, which includes the signature RpoN box motif, consists of a poorly structured N-terminal tail followed by an unpredicted helix-turn-helix motif. Surprisingly, this domain is structurally similar to the σ^3 and σ^4 domains of σ^{70}-type proteins, despite their low sequence homology. Structural homology with other DNA-binding proteins suggests how this protein may bind the -24 promoter element. Because of the high sequence conservation of this region among σ^{54} proteins (Figure 1), it is likely that this domain has a similar fold in all species, including the most highly studied σ54 proteins from *E.coli* and *Klebsiella pneumonia*.

MATERIALS AND METHODS

Sequence analyses

σ^{54} protein sequences from a variety of species were downloaded from the NCBI website (http://www.ncbi.nlm.nih.gov/) and then aligned using the ClustalW (24) webserver at the European Bioinformatics Institute (http://www.ebi.ac.uk/clustalw/). Secondary structure prediction for the *A. aeolicus* σ^{54} protein sequence was performed using the neural network Jnet (25) via the Jpred webserver (26; http://www.compbio.dundee.ac.uk/~www-jpred/).

Plasmid construction

Six unique constructs of the C-terminal region of σ^{54} from *A. aeolicus* were cloned by PCR with primers (Operon) containing NdeI and BamHI restriction sites at the 5'-end and 3'-end, respectively. The

fragments were amplified from a plasmid containing the full-length σ^{54} gene (courteously provided by Gongyi Zhang from the National Jewish Medical Center, Denver CO). The amplified fragments were digested with NdeI and BamHI, purified by gel electrophorsis, and ligated into the pET21a expression vector.

Protein preparation

σ^{54}C protein was expressed using *E. coli* BL21 (DE3) with Rosetta.pLysS plasmid. Cells were first grown at 37°C in two liters of LB media. At OD_{600nm} ~ 0.6, cells were exchanged (27) into one liter of M9 minimal media containing 1 g of ^{15}N-NH_4Cl and/or 2 g of $^{13}C_6$-glucose (Cambridge Isotope Laboratory). Four hours after induction with IPTG, cells were harvested by centrifugation and resuspended into 50 mM Hepes (pH 6.9), 50 mM NaCl, 1 mM EDTA, and 0.1 mM PMSF. σ^{54}C protein was purified by ion exchange chromatography (Heparin, Amersham Biosciences) and by size exclusion chromatography (Superdex 75, Amersham Biosciences). For NMR spectroscopy, σ^{54}C protein was concentrated to 0.8 mM in 50 mM Hepes (pH 6.9), 250 mM NaCl, 1 mM EDTA, 10% ^2H_2O, and 1X Complete Protease Inhibitor Cocktail (Roche Applied Science).

Resonance assigments

NMR data were collected at 298 K on a Bruker DRX 600 MHz or a Bruker spectrometer. All data were processed with NMRPipe (28) and analyzed with NMRView (29). Small amounts of proteolysis were detected by new peaks in the ^{15}N-HSQC spectrum after five days at 298K. Therefore, a fresh protein sample was used for each three-dimensional spectrum.

Backbone assignments were obtained using the 3D triple resonance experiments HNCA, CBCA(CO)NH, and ^{15}N-NOESY-HSQC. Sidechain assignments were obtained from the 3D experiments ^{15}N-TOCSY-HSQC, C(CO)NH, and HCCH-TOCSY, as well as the 2D experiment proton DQF-COSY for the aromatic proton assignments.

Distance restraints

Backbone dihedral restraints were obtained from the backbone chemical shifts (HN, HA, CA, and CB) using the program TALOS (30). $^3J_{HNHA}$-coupling constants were obtained from an HNHA spectrum (31). NOE distance constraints were derived by manual assignment of crosspeaks in a three-dimensional ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC, both with 100 ms mixing times. Hydrogen bond restraints were defined from slowly-exchanging amide protons. The protein was exchanged into 95% D_2O using an Amicon Ultra centrifugal concentrator (5 kDa cutoff, Millipore). Amide protons with strong crosspeak intensities in the ^{15}N-HSQC after 3 hours at 298 K were identified as significantly protected. Stereospecific assignments of the methyl groups of valine and leucine residues were obtained from a ^{13}C-HMQC spectrum of a 10% ^{13}C-labeled protein sample as described in (32). Only residues with totally unambiguous methyl assignments were used for stereospecific assignments.

Structure calculations

Structure calculations were performed with the program DYANA (33). Initial structures were calculated using only unambiguous NOESY peak assignments. During the later stages of assignment (backbone RMSD < 1.5 for residues 21-67), the structure with the lowest energy was used to filter possible assignments based on distances. The program MOLMOL (34) was used to analyze molecules throughout the assignment process. Structure statistics, including PROCHECK analysis (35), are summarized in Table 1.

Miscellaneous

All figures were made with Pymol (http://www.pymol.org). The sequence alignments were plotted with TexShade (36).

RESULTS

Sequence analysis and protein construct design

To help identify a subdomain of the σ^{54} protein suitable for structure determination by NMR, the *A. aeolicus* σ^{54} protein sequence was aligned with 38 other σ^{54} proteins using the program ClustalW (24). This alignment, which includes σ^{54} proteins from 11 of the most studied species, is shown in Figure 1. Visual examination of the sequence alignment indicated that the C-terminal end (approximately 288-398 in *A. aeolicus* and 360-477 in *E. coli*) was the most conserved region. This fragment of σ^{54} has been shown to be sufficient for binding to double-stranded DNA (19)

Using the secondary structure prediction of the *A. aeolicus* σ^{54} protein as a guide, six different constructs of this region were cloned and expressed in *E. coli*. Constructs were created using PCR with three different N-terminal primers and two different C-terminal primers, shown in Figure 1. Only one of the six fragments expressed in *E. coli* with high-levels of soluble protein. This fragment, shown with '**' above the primers in Figure 1 and in Figure 2B, starts at Thr323, just upstream of a pair of conserved phenylalanines (329, 330) and ends at glycine 389, just downstream of the highly conserved RpoN box motif (377-386; boxed in Figure 1). For this part of σ^{54}, the *A. aeolicus* protein sequence is 36% identical to the *E. coli* protein sequence. Here, we present the solution structure of this 67-residue (7.9 kDa) fragment, which we refer to as 'σ^{54}C.'

Structure determination

The σ^{54}C fragment gave an excellent ^{1}H-^{15}N HSQC spectrum with highly dispersed crosspeaks (Supplemental Figure 1), indicating that the majority of the protein fragment was stably folded. The ^{1}H-^{15}N HSQC spectrum contained crosspeaks for all residues except for the first three N-terminal residues (Thr323, Tyr324, Ser325) and the single proline (Pro 359). ^{1}H, ^{15}N and ^{13}C resonance assignments for the σ^{54}C domain were made using standard 3D NMR techniques (37). In total, the chemical shifts for 97% of the backbone atoms (N, NH, CA, HA) and 94% of sidechain atoms (aliphatic/aromatic C and H) were assigned.

Even before we started calculating structures, three properties of the NMR data indicated that the N-terminal portion of σ^{54}C from Leu326 to Thr339 was poorly structured and dynamic on a time scale of microseconds to seconds. First, in all 2D and 3D spectra, diagonal and crosspeak intensities for atoms in this N-terminal region were significantly weaker than for atoms in other parts of the protein. For example, in the ^{1}H-^{15}N HSQC the average relative peak height for residues Leu326 to Thr339 is 24 % less than the average intensity for residues Gln340 to Gly389 (last row of Figure 2B). Second, chemical shift index (38) for this region indicated little secondary structure beyond Leu338 (first two rows of Figure 2B). Third, only a small number of interresidue NOE crosspeaks were present in the 3D ^{13}C-NOESY-HSQC for this region of the protein.

Nine hundred fifty-six unique distance restraints (Table 1) were obtained from the 3D ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC spectra. An additionally 108 φ/ϕ torsion angle restraints were calculated from HN, HA, CA, and CB chemical shift values (30) and from $^{3}J_{HNHA}$-coupling constants (31) (Table 1). These 1064 restraints (16 restraints per residue) were used to calculate 150 structures. The 20 structures with the lowest energy were chosen to represent the structure of the C-terminal domain of the σ^{54} protein from *A. aeolicus* (Figure 2A).

Structure of the C-terminal domain o f σ^{54}

For the well-structured region (Gln340 to Leu388) of σ^{54}C, the superposition of the final 20 structures has an r.m.s. deviation from the mean structure of 0.56 ± 0.07 Å for the backbone atoms and 1.24 ± 0.09 Å for all heavy atoms. Overall these structures have good statistics and geometry, as summarized in Table 1. This C-terminal portion of σ^{54}C folds into a compact domain consisting of three

α-helices (Figure 2). Helix 1 (Gln340 to Asn353) is followed by a long loop, which starts at the conserved Glu354 and ends at the conserved Ser361. Helix 2 (Asp362 to Leu369) and helix 3 (Arg379 to Leu388), which includes the highly conserved RpoN box sequence, form a helix-turn-helix motif that is commonly found in many DNA-binding proteins (reviewed in 39).

A color-coded image representing the electrostatic surface of σ^{54}C (residues Gln340 to Leu388) shows that the molecule has a striking polarity. It has a positively charged face primarily composed of the loop in the helix-turn-helix motif (loop 2) and the top of helix 3 (Figure 3C, right). The other side of the molecular is negatively charged, due to negative residues in loop 1 and on helix 1 (Supplemental Figure 2).

A search for proteins structurally homologous to σ^{54}C (residues Gln340 to Leu388) using the DALI sever identified many DNA-binding proteins. Residues 77 to 127 of the human PAX6 homeodomain (PDB accession number 6PAX; 40) had the highest Z score of 5.2 (r.m.s.d 2.1 Å). The N-terminal DNA-binding domain of the leucine-responsive regulatory protein, LrpA, from *Pyrococcus furiosus* (PDB accession number 1I1G; 41) had the second highest Z score of 4.7 (r.m.s.d. 1.9 Å). The DALI results are summarized in Supplemental Table 1.

Interestingly, the σ^{54}C domain has structural homology to both σ^3 (regions 3.0-3.1) and σ4 (regions 4.1 - 4.2) of σ^{70}-type factors. σ^3 of σA (residues 279 to 326) from *Thermus aquaticus* (PDB accession number 1KU2; 15) had the tenth highest Z score of 4.1. σ^4 of σ^{28} (residues 186-224) from *A. aeolicus* (PDB accession number 1RP3; 14) had a lower Z score of 3.8 but aligns slightly better with the σ^{54}C structure (r.m.s.d. 1.9 Å) than the σ^3 fragment of σ^A (r.m.s.d. of 2.5 Å). Structural alignments with σ^4 and σ^3 are shown in Figure 3B, left and right, respectively.

As described above, the N-terminal segment from Ser325 to Thr339 is poorly structured and probably dynamic on a microsecond to second time scale. This type of dynamics causes line broadening, thus making less structural information available for this region. The poor chemical shift dispersion for backbone and sidechain atoms (top 2 rows in Figure 2B) and the low number of medium and long-range NOE crosspeaks available made it possible to assign only ten long-range NOE crosspeaks availabe for this region. These crosspeaks indicate that Ala335 and Leu338 interact with the N-terminal end of helix 1 primarily via Leu343 and with the loop between helices 2 and 3 primarily via Phe374. For example, the aliphatic hydrogens on Leu338 interact with the amide and aliphatic hydrogens on Leu343 and with the aromatic hydrogens on Phe374.

Structural role for conserved residues

The hydrophobic core connecting the three helices is composed of highly conserved hydrophobic residues. For example, Ile347 in helix 1 makes van der Waal interactions with Ile365 in helix 2 and Val381 in helix 3. All three of these residues are isoleucine, leucine, or valine in all 38 σ^{54} protein sequences examined (11 of these sequences are shown in Figure 1). Other residues making crucial van der Waals interactions between the helices are Leu343, Met344, and Ile350 in helix 1; Ala366, Ile368, and Leu369 in helix 2; and, Tyr384 and Leu388 in helix 3.

Many of the residues creating the charge on the protein surface are conserved among σ^{54} proteins. The highly conserved RpoN box (Ala378 to Glu386) forms the majority of helix 3, the second helix in the helix-turn-helix motif. This helix is amphipathic, with the hydrophobic face packing against the hydrophobic core described above. The hydrophilic side is surface exposed. This helix is primarily positively charged due to the sidechains Arg378, Arg379, Lys383 and Arg385, but the C-terminal part of the helix has negative charge due to the sidechain of Glu386 (Figure 3C; right).

DISCUSSION

When fragments of σ^{70} were crystallized, they were found to be small helical domains (15). The subsequent structure of σ^{70} (regions 2-4) in complex with core RNA polymerase, showed that these domains are modular, connected by flexible linkers, which interact with other parts of polymerase and DNA (16, 42). Sequence analysis and biochemical experiments suggest that σ^{70} may also be comprised of

such domains (43, 44). Here we present the first structure of a σ^{54} domain, the C-terminal segment of the protein, which is critical for DNA binding and contains the signature 'RpoN box' sequence (Figure 1).

Residues Gln340 to Leu388 of *A. aeolicus* σ^{54} form a compact, three-helix domain, which includes a helix-turn-helix motif. A search of the protein data bank using the DALI server found that this region of σ^{54} is structural similar to many DNA-binding domains (Supplemental Table 2). Of the ten structures with the highest Z scores, seven of them are known to interact with DNA. For two of these domains, the C-terminal part of the PAX6 homeodomain (PDB accession number 6PAX; 40) and the N-terminal domain of the methicillin repressor protein (PDB accession number 1SAX; 45), the structures of the protein-DNA complexes are known (40, 46). These structures show that the proteins interact primarily with the major groove of DNA via the second helix of the helix-turn-helix motif, termed the 'recognition helix.' The equivalent helix in the σ^{54}C domain is created by residues of the highly conserved RpoN box motif (helix 3; Figure 2B and 3A). Thus, residues in the RpoN box are structurally poised to interact with the major groove of the σ^{54} promoter element.

Consistent with interpreting structure of the σ^{54}C domain as a DNA-binding domain, previous footprinting and gel shift assays have shown that the region near the RpoN box is critical for σ^{54} promoter recognition and DNA binding (47, 19, 48). Mutations in the RpoN box sequence are far more damaging to DNA binding than mutations in other parts of *E.coli* σ^{54} (21, 48). All five mutations made in the RpoN box sequence had reduced binding to both duplex and fork junction promoter DNA, with four mutant proteins (*K. pneumonaie* R455A, R456A, Y461A, and R462A) showing $\leq$ 20% DNA binding of wild-type σ54 protein. In addition, these four mutant proteins had dramatically reduced transcription activity *in vitro* with $\leq$ 30% the activity of the wild-type protein. Thus, these residues on the RpoN box are crucial for DNA-binding and σ^{54}-dependent transcription. The corresponding residues in *A. aeolicus* σ^{54} are shown as sticks in Figure 3A. Four of these residues (Arg378, Arg379, Lys383, and Arg385) lie on the hydrophillic face of the 'recognition helix,' suggesting that this face of the helix interacts with DNA.

Additional support for the role of this domain in DNA-binding comes from the work of Cannon et al. (47). They found that removing the last 53 residues of *K. pneumonaie* σ54 (Lys425 to Val477; *A. aeolicus* Lys346 to Ile398) eliminated DNAse I and *o*CuP footprints of σ^{54} and reduced σ54-holoenzyme binding to 5-BrU substituted DNA by 85%. This truncation cuts the protein in the middle of helix 1 of the σ^{54}C domain, removing loop 2 and the helix-turn-helix motif.

Overall, the C-terminal domain of σ54 is more structurally similar to domain σ^3 than to domain σ^4 of σ^{70}. σ^3 and the σ^{54}C are about the same size (55 and 59 residues, respectively), and all three helices are about the same length (Figure 3B, right). On the other hand, σ^4 has an extra helix at the N-terminal end, and the 'recognition helix' is twice as long as the corresponding helix in the σ^{54}C domain (Figure 3B, left). In terms of surface charge distribution, however, the σ^{54}C is more similar to σ^4 than to σ3 (Figure 3C). Both σ4 and σ^{54}C have a positively charged surface created by sidechains on the loop and the recognition helix of the helix-turn-helix motif (Figure 3C, left and middle). For σ^4, this surface interacts specifically with the major groove of the –35 promoter element (15, 13). For σ^{54}C, this surface is created by the residues of the conserved RpoN box.

Previous biochemical studies also suggest that the C-terminal domain of σ^{54} is more functionally analogous to σ^4 domain than to σ^3 domain of σ^{70}-type factors. σ^3 of σ^{70} proteins interacts with the extended -10 promoter element near the site of melting (42, 49); whereas, the major role of σ^4 of σ^{70} proteins is to bind to the -35 promoter element (15, 13). By tethering a DNA cleavage reagent FeBABE to charged residues in the RpoN box of *K. pneumoniae* σ^{54}, Burrows et al. (50) showed that residues in the RpoN box are near the -24 promoter element, not the -12 promoter element where melting occurs. For example, with the FeBABE tethered to residue E463C (*A. aeolicus* Glu386), strong cleavage was seen between the -30 and -25 nucleotides of the template DNA strand.

The N-terminal end of σ^{54}C domain, residues Ser325 to Thr339, is not well-ordered and probably dynamic on the microsecond to second time scale. These residues could be a part of a linker region that connects two independently folded domains of σ^{54}, as seen with σ^{70}-type factors (15, 16). Previous

limited proteolysis experiments on the full-length σ^{54} did not identify this segment as proteolytically sensitive (51). Thus, additionally proteolysis studies and NMR analyses on larger fragments of the *A. aeolicus* σ^{54} are required to determine if this region (Ser325 to Thr339) is still poorly structured in the presence of the other domains and RNAP.

The C-terminal domain of σ^{54} presented in this paper provides an interesting structural connection between σ^{54} and σ^{70}. This structural similarity of σ^{54}C to domains of σ^{70}-type sigma factors is surprising (Figure 3B) because these regions have poor sequence homology. Residues Thr323 to Gly389 of σ^{54}C have 14% identity, 25% similarity with residues 186 to 326 of region 4 of σA from *T. aquiticus* and 15% identity, 26% similarity with residues 279 to 326 of region 3 of σ28 from *A. aeolicus*. A low-resolution structure of the full-length σ^{54} protein suggests that other domains of σ^{54} may also be more similar to domains of σ^{70} than previously predicted by only sequence homology (22). High-resolution structures of these other domains are needed to confirm this hypothesis.

CONCLUSION

Our understanding of the details of σ^{54}-dependent transcription initiation, as opposed to σ^{70}-dependent transcription, has been hindered by the lack of structural information. The high-resolution structure of the C-terminal domain of σ^{54} presented in this paper is the first step towards our long-term goal of obtaining high-resolution solution structures of each domain of σ^{54} alone and in complex with DNA. Further, despite the low sequence homology between this region of σ^{54} and the σ^{3} and σ^{4} domains of σ^{70}-type proteins, the domains have high structural similarity (Figure 3B). Thus, other domains of σ^{54} and σ^{70} proteins may be more structurally similar than previously predicted based on protein sequence alone. This would not be too surprising given that both types of factors use the same core RNAP to regulate and initiate promoter specific transcription.

FOOTNOTES

The atomic coordinates, chemical shift assignments, and the distance and coupling restraints were deposited in the RCSB PDB under the accession code 2AHQ. We thank Prof. Sydney Kustu for all her advice and Steven M. Damo for his computer assistance. This work was supported by a University of California Fellowship (MD) and ---.

[1]The abbreviations used are: RNAP, RNA-polymerase; National Center for Biotechnology Information; LB, Luria-Bertani; IPTG, isopropyl-beta-D-thiogalactopyranoside; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; HSQC, heteronuclear single quantum coherence spectroscopy; *o*CuP, 1,10-phenanthroline-copper; FeBABE, *p*-bromoacetamidobenzyl-EDTA-Fe; RCSB PDB, Research Collaboratory for Structural Bioinformatics Protein Databank

REFERENCES

1. Burgess, R. R., Travers, A. A., Dunn, J. J., and Bautz, E. K. (1969) Nature 221, 43-46
2. Hirschman, J., Wong, P. K., Sei, K., Keener, J., and Kustu, S. (1985) Proc Natl Acad Sci U S A 82, 7525-7529
3. Wosten, M. M. (1998) FEMS Microbiol Rev 22, 127-150
4. Lonetto, M., Gribskov, M., and Gross, C. A. (1992) J Bacteriol 174, 3843-3849
5. Paget, M. S., and Helmann, J. D. (2003) Genome Biol 4, 203
6. Chilcott, G. S., and Hughes, K. T. (2000) Microbiol Mol Biol Rev 64, 694-708
7. Buck, M., Gallegos, M. T., Studholme, D. J., Guo, Y., and Gralla, J. D. (2000) J Bacteriol 182, 4129-4136
8. Fisher, M. A., Grimm, D., Henion, A. K., Elias, A. F., Stewart, P. E., Rosa, P. A., and Gherardini, F. C. (2005) Proc Natl Acad Sci U S A 102, 5162-5167

9. Xu, H., and Hoover, T. R. (2001) Curr Opin Microbiol 4, 138-144
10. Wedel, A., Weiss, D. S., Popham, D., Droge, P., and Kustu, S. (1990) Science 248, 486-490
11. Malhotra, A., Severinova, E., and Darst, S. A. (1996) Cell 87, 127-136
12. Campbell, E. A., Tupy, J. L., Gruber, T. M., Wang, S., Sharp, M. M., Gross, C. A., and Darst, S. A. (2003) Mol Cell 11, 1067-1078
13. Jain, D., Nickels, B. E., Sun, L., Hochschild, A., and Darst, S. A. (2004) Mol Cell 13, 45-53
14. Sorenson, M. K., Ray, S. S., and Darst, S. A. (2004) Mol Cell 14, 127-138
15. Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L., and Darst, S. A. (2002) Mol Cell 9, 527-539
16. Murakami, K. S., Masuda, S., and Darst, S. A. (2002) Science 296, 1280-1284
17. Murakami, K. S., and Darst, S. A. (2003) Curr Opin Struct Biol 13, 31-39
18. Wong, C., Tintut, Y., and Gralla, J. D. (1994) J Mol Biol 236, 81-90
19. Cannon, W. V., Chaney, M. K., Wang, X., and Buck, M. (1997) Proc Natl Acad Sci U S A 94, 5006-5011
20. Bordes, P., Wigneshweraraj, S. R., Schumacher, J., Zhang, X., Chaney, M., and Buck, M. (2003) Proc Natl Acad Sci U S A 100, 2278-2283
21. Taylor, M., Butler, R., Chambers, S., Casimiro, M., Badii, F., and Merrick, M. (1996) Mol Microbiol 22, 1045-1054
22. Svergun, D. I., Malfois, M., Koch, M. H., Wigneshweraraj, S. R., and Buck, M. (2000) J Biol Chem 275, 4210-4214
23. Rappas, M., Schumacher, J., Beuron, F., Niwa, H., Bordes, P., Wigneshweraraj, S., Keetch, C. A., Robinson, C. V., Buck, M., and Zhang, X. (2005) Science 307, 1972-1975
24. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Nucleic Acids Res 22, 4673-4680
25. Cuff, J. A., and Barton, G. J. (2000) Proteins 40, 502-511
26. Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M., and Barton, G. J. (1998) Bioinformatics 14, 892-893
27. Marley, J., Lu, M., and Bracken, C. (2001) J Biomol NMR 20, 71-75
28. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) J Biomol NMR 6, 277-293
29. Johnson, B. A., and Blevins, B. A. (1994) Journal Biolmolecular NMR 4
30. Cornilescu, G., Delaglio, F., and Bax, A. (1999) J Biomol NMR 13, 289-302
31. Kuboniwa, H., Grzesiek, S., Delaglio, F., and Bax, A. (1994) J Biomol NMR 4, 871-878
32. Neri, D., Szyperski, T., Otting, G., Senn, H., and Wuthrich, K. (1989) Biochemistry 28, 7510-7516
33. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) J Mol Biol 273, 283-298
34. Koradi, R., Billeter, M., and Wuthrich, K. (1996) J Mol Graph 14, 51-55, 29-32
35. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Cryst. 26, 283-291
36. Beitz, E. (2000) Bioinformatics 16, 135-139
37. Cavanagh, J., Wayne, J. F. I., Palmer, A. G., and Nicholas, J. S. (1996) P*rotein NMR Spectroscopy: Principles and Practice*, Academic Press, San Diego
38. Spera, S., and Bax, A. (1991) J. Am. Chem. Soc. 113, 5490-5492
39. Wintjens, R., and Rooman, M. (1996) J Mol Biol 262, 294-313
40. Xu, H. E., Rould, M. A., Xu, W., Epstein, J. A., Maas, R. L., and Pabo, C. O. (1999) Genes Dev 13, 1263-1275
41. Leonard, P. M., Smits, S. H., Sedelnikova, S. E., Brinkman, A. B., de Vos, W. M., van der Oost, J., Rice, D. W., and Rafferty, J. B. (2001) Embo J 20, 990-997
42. Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., and Darst, S. A. (2002) Science 296, 1285-1290
43. Sasse-Dwight, S., and Gralla, J. D. (1990) Cell 62, 945-954
44. Missailidis, S., Cannon, W. V., Drake, A., Wang, X. Y., and Buck, M. (1997) Mol Microbiol 24, 653-664
45. Garcia-Castellanos, R., Marrero, A., Mallorqui-Fernandez, G., Potempa, J., Coll, M., and Gomis-

Ruth, F. X. (2003) J Biol Chem 278, 39897-39905
46. Garcia-Castellanos, R., Mallorqui-Fernandez, G., Marrero, A., Potempa, J., Coll, M., and Gomis-Ruth, F. X. (2004) J Biol Chem 279, 17888-17896
47. Cannon, W., Missailidis, S., Smith, C., Cottier, A., Austin, S., Moore, M., and Buck, M. (1995) J Mol Biol 248, 781-803
48. Wang, L., and Gralla, J. D. (2001) J Biol Chem 276, 8979-8986
49. Voskuil, M. I., and Chambliss, G. H. (2002) J Mol Biol 322, 521-532
50. Burrows, P. C., Severinov, K., Ishihama, A., Buck, M., and Wigneshweraraj, S. R. (2003) J Biol Chem 278, 29728-29743
51. Cannon, W., Claverie-Martin, F., Austin, S., and Buck, M. (1994) Mol Microbiol 11, 227-236
52. Pohl, E., Haller, J. C., Mijovilovich, A., Meyer-Klaucke, W., Garman, E., and Vasil, M. L. (2003) Mol Microbiol 47, 903-915
53. Mer, G., Bochkarev, A., Gupta, R., Bochkareva, E., Frappier, L., Ingles, C. J., Edwards, A. M., and Chazin, W. J. (2000) Cell 103, 449-456
54. Nichols, M. D., DeAngelis, K., Keck, J. L., and Berger, J. M. (1999) Embo J 18, 6177-6188

FIGURE LEGENDS

Figure 1. Amino acid sequence alignment for the C-terminal end of the σ^{54} protein from eleven different species. Residues conserved in all proteins are highlighted in purple; residues conserved in ≥ 50% of the species are highlighted in blue; and chemically similar residues are highlighted in magenta. The consensus sequence is shown at the bottom. The highly conserved RpoN box motif is framed in black. The positions of the five PCR primers (three N-terminal primers and two C-terminal primers) used to create six unique protein constructs of this region are shown as black arrows. Only one of the six fragments expressed in *E. coli* with high-levels of soluble protein. This fragment, shown with '**' above the primers, was used for further structural studies.

Figure 2. (A) Stereoview of the backbone (N, CA, C') of 20 superimposed NMR structures with the lowest restraint violations for the C-terminal fragment (Thr323 to Gly389) of *A. aeolicus* σ^{54}. Only the well-structured region, Ala335 to Gly389, (E. coli 407-466) is shown. Helix 3 that is created by the RpoN box is in the front in dark blue.(B) Protein sequence alignment of the *A. aeolicus* and *E. coli* σ54C fragment. Conserved residues are highlighted in blue; the highly conserved RpoN box is underlined in red. α-helix secondary structure is indicated above in red. Below the sequence alignment (from top to bottom) are the CSI analysis of the CA atoms, CSI analysis of the HA atoms, schematic representation of the sequential i, i+1 NOE connectivities involving NH atoms, schematic representation of the medium range i, i+3 NOE connectivities involving HA, HN atoms, respectively. For the NOE plots, the bar is on the residue containing atom i and the height of the bar is inversely proportional to the maximum distance between the atoms.

Figure 3. (A) Ribbon diagram of the σ^{54}C domain (residues Ala335 to Gly389). Sidechains for residues that were previously shown to be near the -24 promoter element (50) and/or significantly decrease DNA-binding when mutated (48) are represented as sticks and labeled (except for Tyr384 which is hidden behind Lys383). All of these residues are on helix 3 (dark blue) created by the highly conserved RpoN box, except for Met343 (Arg421 in *E. coli*), which is on helix 1. (B) Superposition of the CA backbones of the σ^{54}C fragment (blue) with the σ^4 of the σ^{70}-type protein σ28 (magenta; residues 175-234; PDB accession number 1RP3) (left) and the σ^3 of the σ^{70}-type protein σA (green; residues 279-326; PDB accession number 1KU2) (right). In both superposition, the beginning and the end of the RpoN box on σ^{54} are shown as blue spheres. (C) Surface electrostatic potential (blue positive; red negative) of σ^{54}C domain (left), σ^4 of the σ^{70}-type protein σ^{28} (middle), and σ^3 of the σ^{70}-type protein σA (right). Highly conserved residues are labeled on σ^{54}C. In all images, the proteins are oriented as in (A).

Supplemental Figure 1. Labeled ^{1}H-^{15}N HSQC spectrum of the σ^{54}C protein fragment from *A. aeolicus*. All residues of the fragment are visible, except for the first three N-terminal residues (Thr323, Tyr324, Ser325) and the one proline residue (Pro 359).

Supplemental Figure 2. A. Surface electrostatic potential (blue positive; red negative) of the σ^{54}C domain showing the polarity of the domain. On the left, the protein is oriented as in Figures 2A and 3A with the RpoN box helix in the front; on the right, the protein has been rotated 180° about the y-axis (right).

Figure 1

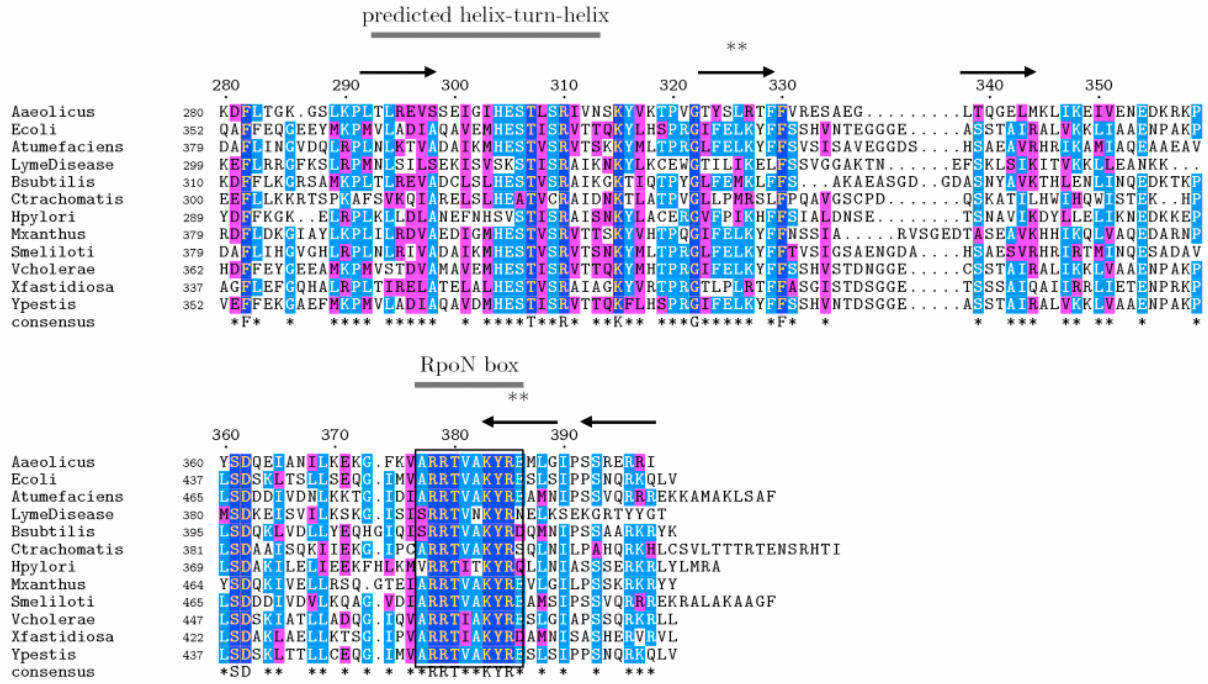
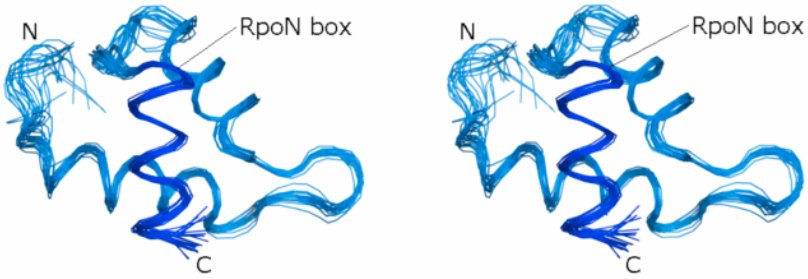

predicted helix-turn-helix
RpoN box
Aaeolicus
Ecoli
Atumefaciens
LymeDisease
Bsubtilis
Ctrachomatis
Hpylori
Mxanthus
Smeliloti
Vcholerae
Xfastidiosa
Ypestis
consensus

Figure 2

A.


N
RpoN box
C
N
RpoN box
C

B.

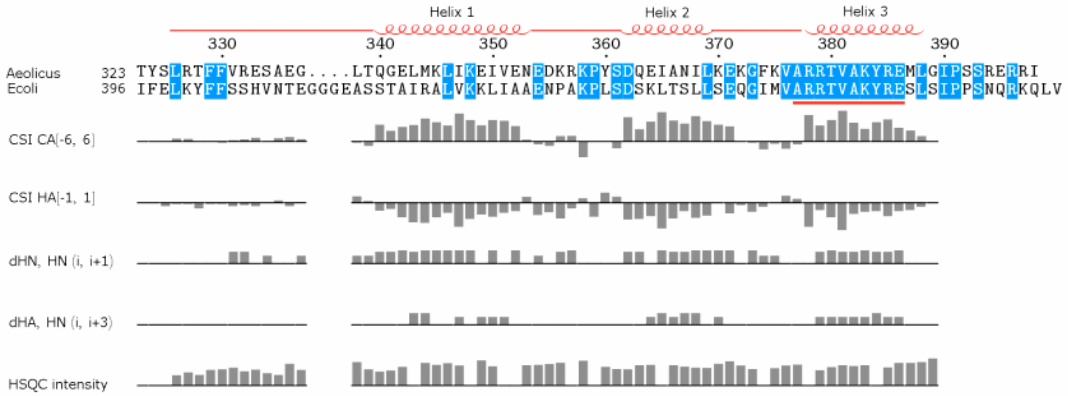
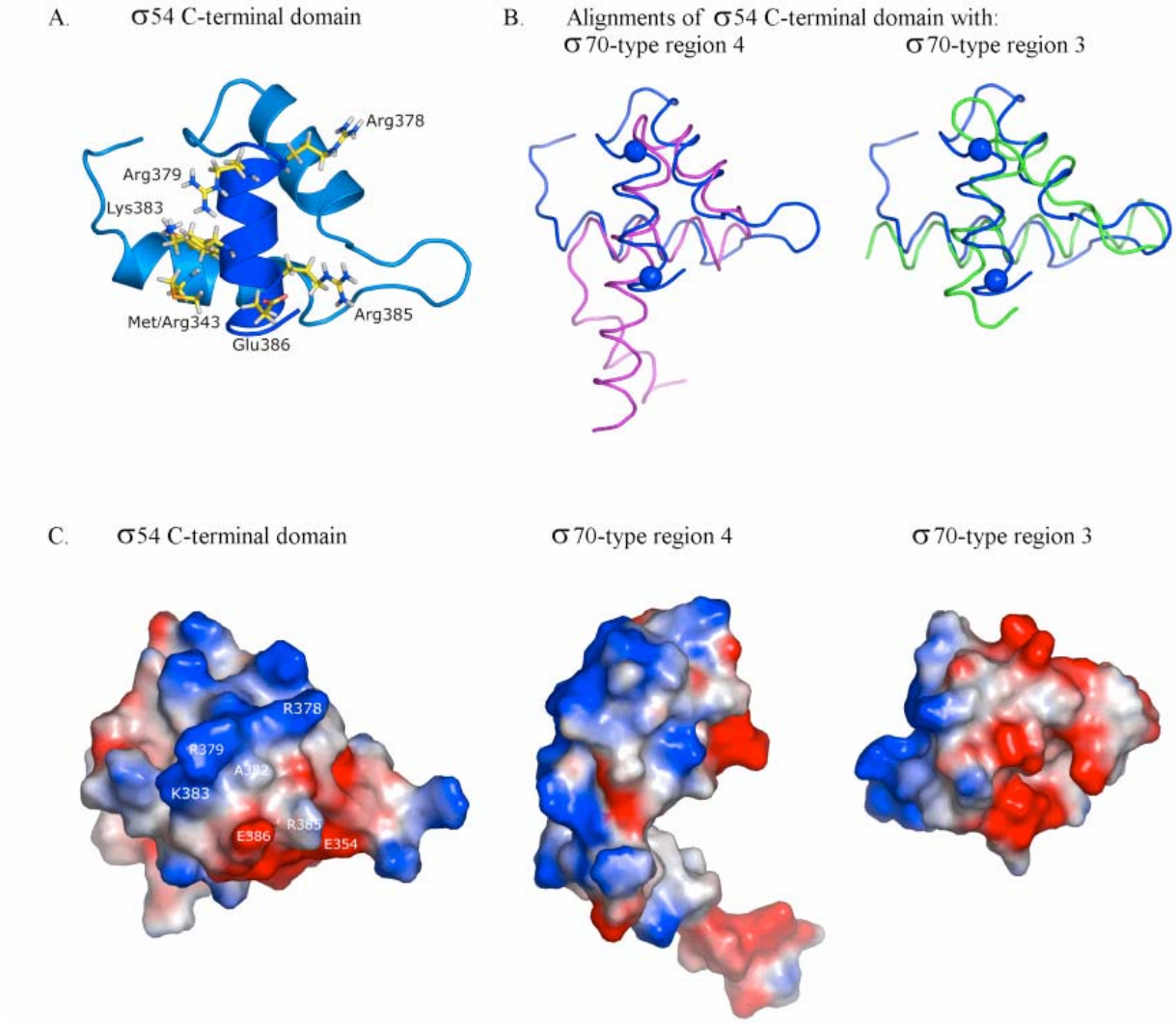
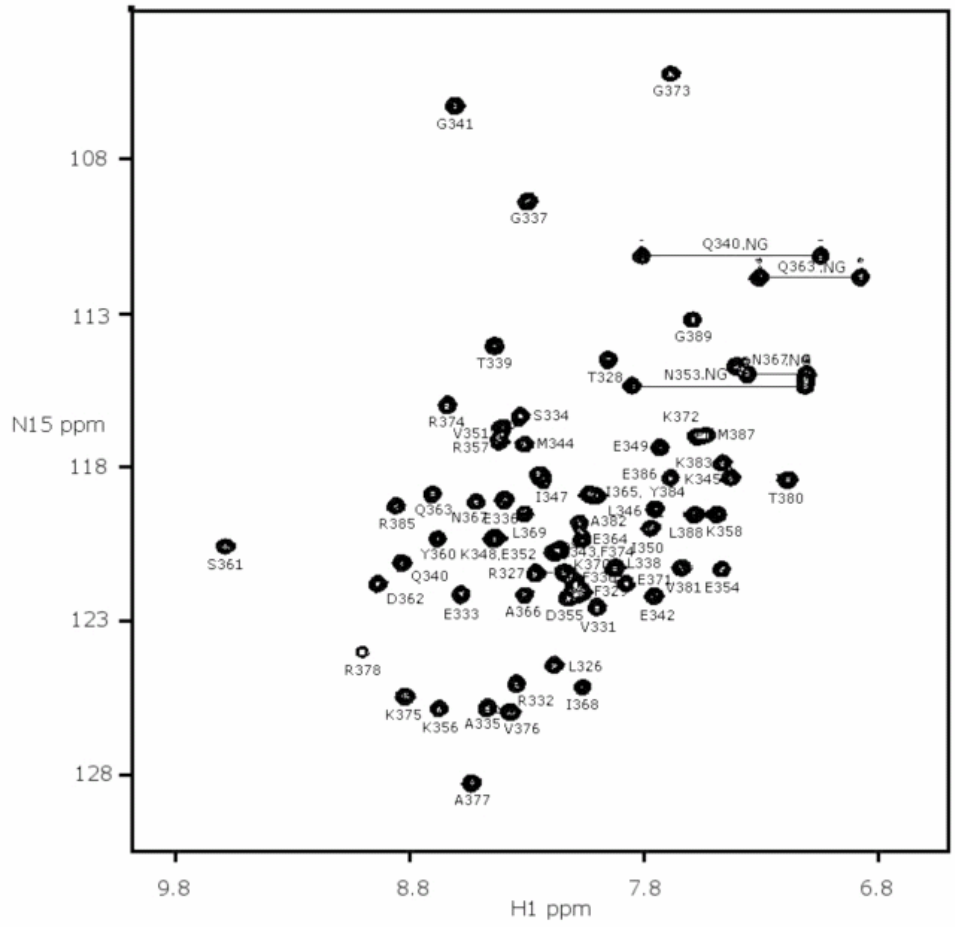

Helix 1
Helix 2
Helix 3
330
340
350
360
370
380
390
Aeolicus 323 TYSLRTFFVRESAEG....LTQGELMKLIKEIVENPDKRKPYSDQEIANILKPKGFKVARRTVAKYREMLGIPSSRERRI
Ecoli 396 IFELKYFFSSHVNTEGGGEASSTAIRALVKKLIAAPNPAKPLSDSKLTSLLSPQGIMVARRTVAKYRESLSIPPSNQRKQLV
CSI CA[-6, 6]
CSI HA[-1, 1]
dHN, HN (i, i+1)
dHA, HN (i, i+3)
HSQC intensity

Figure 3:


A.
σ54 C-terminal domain
B.
Alignments of σ54 C-terminal domain with:
σ70-type region 4
σ70-type region 3
Arg378
Arg379
Lys383
Met/Arg343
Glu386
Arg385
C.
σ54 C-terminal domain
σ70-type region 4
σ70-type region 3
R378
R379
K383
E386
R385
E354

Supplemental Figure 1:



Supplemental Figure 2:

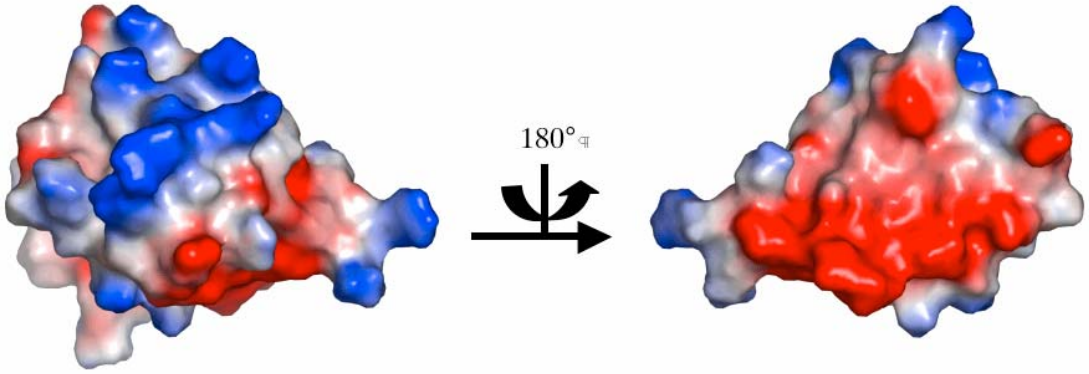


Table 1. Experimental restraints and structural statistics for the C-terminal domain of σ^{54}

Restraints	1064
Total Distance	956
Intraresidue	458
Sequential (|i-j| = 1)	201
Mediume range ($2 \geq$ |i-j| ≤ 5)	128
Long range (|i-j| ≥ 5)	169
Dihedral φ/ϕ angle restraints	
TALOS prediction	64
$^{3}J_{HNHA}$ coupling	44
Total restraints	1064
Residual violations[a]	
Number of Distance violations ≥ 0.3 Å	2.74 ± 1.42
Number of dihedral angle violations $\geq 5°$	0 ± 0
Number of $^{3}J_{HNHA}$ coupling violations ≥ 0.5 Hz	0 ± 0
Number of van der waals violations ≥ 0.2 Å	1.1 ± 0.91
Average DYANA target function (Å^2)	2.60 ± 0.33
Coordinate precision (Å)[b]	
Backbone atoms in well-structured region (residues Gln340 to Leu388)	0.56 ± 0.07
Heavy atoms in well-structured region (residues Gln340 to Leu388)	1.24 ± 0.09
Procheck statistics[c]	
Residues in the most favored region (%)	75.8
Residues in additional allowed regions (%)	20.8
Residues in generously allowed regions (%)	3.3
Residues in disallowed regions (%)	0.2

a) The average number of distance and angle violations for each of the 20 final conformers is reported.

b) R.M.S.D. values are calculated from the mean structure.
c) Structures were not energy minimized.

Supplemental Table 1. Summary of structural homology search[a] for the C-terminal domain of σ^{54}

Homologous Protein (residues)	**Z Score (R.M.S.D. [Å])**	**DNA binding domain?**	**DNA Binding Surface**	**PDB ID (Reference)**
C-terminal domain of the human homeobox Pax6 protein (77-108;114-127)	5.2 (2.1)	yes	'recognition helix' of HTH motif	6PAX (40)
N-terminal domain of *Pyrococcus furiosus* leucine-respone regulator (2-16;18-42)	4.7 (1.9)	yes	putatively 'recognition helix' of HTH motif	1ILG (41)
N-terminal domain of *Pseudomonas aeruginosa* ferric uptake regulator (15-61)	4.4 (2.0)	yes	putatively 'recognition helix' of HTH motif	1MZB (52)
C-terminal domain of human replication protein A (205-249)	4.4 (2.3)	no	-	1DPU (53)
N-terminal domain of *Methanococcus jannaschii* DNA topoisomerase VI A subunit	4.3 (2.5)	yes	putatively 'recognition helix' of HTH motif	1D3Y (54)
σ^3 domain of *Thermus aquaticus* σ^A	4.1 (2.5)	no	-	1KU2 (15)
σ^4 domain of *A. aeolicus* σ^{28}	3.8 (1.9)		'recognition helix' of HTH motif	1RP3 (14)

a) The structures with the top five DALI Z scores are listed, as well as the two highest scoring fragments of the σ^{70}-type proteins.