The Structural Basis for Promoter –35 Element Recognition by the Group IV σ Factors

William J. Lane, Seth A. Darst
The Rockefeller University, New York, New York, United States of America

The control of bacterial transcription initiation depends on a primary σ factor for housekeeping functions, as well as alternative σ factors that control regulons in response to environmental stresses. The largest and most diverse subgroup of alternative σ factors, the group IV extracytoplasmic function σ factors, directs the transcription of genes that regulate a wide variety of responses, including envelope stress and pathogenesis. We determined the 2.3-Å resolution crystal structure of the –35 element recognition domain of a group IV σ factor, Escherichia coli σ^35, bound to its consensus –35 element, GGAACCT. Despite similar function and secondary structure, the primary and group IV σ factors recognize their –35 elements using distinct mechanisms. Conserved sequence elements of the σ^35 –35 element induce a DNA geometry characteristic of AA/TT-tract DNA, including a rigid, straight double-helical axis and a narrow minor groove. For this reason, the highly conserved AA in the middle of the GGAACCT motif is essential for –35 element recognition by σ^35, despite the absence of direct protein–DNA interactions with these DNA bases. These principles of σ^35–35 element recognition can be applied to a wide range of other group IV σ factors.

Introduction

Bacterial transcription is driven by the DNA-dependent RNA polymerase (RNAP), comprising five core subunits (σ_2β’βα) plus an initiation-specific σ subunit, which binds to the core RNAP to form the holoenzyme [1–3]. Promoter-specific transcription initiation first requires the formation of a closed complex in which σ domains 2 (σ_2) and 4 (σ_4) bind sequence-specifically to the –10 and –35 promoter DNA elements, respectively [3–5]. Analysis of the available bacterial genomes has revealed great variation in both the number and type of σ factors that each bacterial species possesses [6,7], allowing for promoter-specific transcription of defined regulons.

Most σ factors belong to the σ^70 family, which can be broadly divided into five subgroups [7,8]. The group I (primary) σ factors, such as Escherichia coli (Ec) σ^70 and Thermus aquaticus (Taq) σ^70*; [22]), domain 4 of all primary σs, which contains a helix-turn-helix DNA binding motif, recognizes the 6-base-pair (bp) –35 consensus TTGACA [4,23], while Ec σ^35 is thought to directly recognize the 7-bp –35 element GGAACCT [17]. Taken together, this suggests that the different groups of σ factors share the same general mechanisms of –35 element binding, but that residue changes on the surface of the recognition helix account for differences in promoter specificity. Previous studies have revealed the molecular details of how domain 4 of the group I σ factor Tag σ^35 recognizes its –35 consensus promoter element [4]. To better understand the structural basis for group IV σ factor promoter specificity, we solved the 2.3-Å resolution crystal structure of Ec σ^35 bound to its –35 consensus promoter element. The structure reveals that, despite the structural similarity with Taq σ^35, Ec σ^35 recognizes its –35 element in a distinct manner. Conserved sequence elements of the σ^35 –35

Many of the variable σ^E regulon members are critical for virulence in important pathogens [18–21]. The structure of Ec σ^35 bound to the cytoplasmic portion of its anti-σ RseA revealed that, despite little primary sequence identity, domains 2 and 4 of σ^E (σ^35_2 and σ^35_4, respectively) share striking structural similarity to the corresponding domains of Taq σ^35 (σ^70_2 and σ^70_4; [22]). Domain 4 of all primary σs, which contains a helix-turn-helix DNA binding motif, recognizes the 6-base-pair (bp) –35 consensus TTGACA [4,23], while Ec σ^35 is thought to directly recognize the 7-bp –35 element GGAACCT [17]. Taken together, this suggests that the different groups of σ factors share the same general mechanisms of –35 element binding, but that residue changes on the surface of the recognition helix account for differences in promoter specificity. Previous studies have revealed the molecular details of how domain 4 of the group I σ factor Tag σ^35 recognizes its –35 consensus promoter element [4]. To better understand the structural basis for group IV σ factor promoter specificity, we solved the 2.3-Å resolution crystal structure of Ec σ^35 bound to its –35 consensus promoter element. The structure reveals that, despite the structural similarity with Taq σ^35, Ec σ^35 recognizes its –35 element in a distinct manner. Conserved sequence elements of the σ^35 –35

Copyright: © 2006 Lane and Darst. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: bp, base pair; Bsu, Bacillus subtilis; Ec, Escherichia coli; ECF, extracytoplasmic function; Mtb, Mycobacterium tuberculosis; Paer, Pseudomonas aeruginosa; Psyr, Pseudomonas syringae; RNAP, RNA polymerase; Scoe, Streptomyces coelicolor; Taq, Thermus aquaticus

* To whom correspondence should be addressed. E-mail: darst@rockefeller.edu

Citation: Lane WJ, Darst SA (2006) The structural basis for promoter –35 element recognition by the group IV σ factors. PLoS Biol 4(9): e269. DOI: 10.1371/journal.pbio.0040269
element, including the most highly conserved 'AA' of the GGAACTT motif, are not involved in direct interactions between the protein and the unique edges of the DNA bases. Instead, these DNA elements induce a specific DNA geometry that is required for E4 binding. Sequence analysis of other group IV rs and their cognate /C0 elements indicates that this principle of /C0 element recognition is a conserved feature of /C0 element recognition by group IV rs factors.

Results

Crystallization and Structure Determination
We performed vapor diffusion crystallization trials with Ec E4 (residues 122 to 191) in complex with DNA fragments corresponding to the Ec /C0 consensus –35 promoter sequence GGAACTT [17]. Thin rectangular crystals grown using a 12-bp DNA fragment (Figure 1A) diffracted to 2.3 Å-resolution (see Materials and Methods and Table 1). The structure was determined by molecular replacement using both a model of Ec /C0 from the Ec /C0/RseA complex structure [22] and the 6-bp –35 element from the Taq /C0/DNA structure [4] in search models. The crystals contained two /C0/DNA complexes per asymmetric unit, with a solvent content of 65%. Iterative model building and crystallographic refinement converged to an R/ /Rfree of 0.241/0.253 (Table 2).

Overall Structure
Two /C0 molecules in the asymmetric unit each bound a separate DNA fragment. As anticipated, the recognition helix of the /C0 helix-turn-helix motif bound in the major groove of the –35 element (Figure 1B). The crystallographically related DNA helices packed head-to-tail, forming a pseudo-continuous double helix with the 1 bp overhangs forming Hooogsten base pairs with the adjacent double helices.

Table 1. Ec /C0/DNA Diffraction Data

| Dataset   | Wavelength (Å) | Resolution (Å) | Number of Reflections (Total/Unique) | Completeness (%) | /α (Å) | Rsym,a (%) |
|-----------|----------------|----------------|---------------------------------------|------------------|--------|------------|
| Nativeb   | 1.0004         | 20–2.3 (2.38–2.30) | 222,494/19,507                        | 97.5 (96.1)      | 13.4 (3.8) | 5.2 (40.2) |

aRsym = σ/|<I>|/σI, where I is observed intensity and <I>| is average intensity obtained from multiple observations of symmetry related reflections. bDataset was collected at the National Synchrotron Light Source Beamline X25.

DOI: 10.1371/journal.pbio.0040269.t001
to −35 and the template strand from −29′ to −32′ [throughout this paper, DNA bases will be numbered as in Figure 3A, where negative numbers denote base pairs upstream of the transcription start site]. Unprimed numbers denote the non-template (top) DNA strand, while primes denote the template (bottom) strand. Specific protein–DNA base interactions occur through direct hydrogen bonds and van der Waals forces (Figures 2 and 3A). In addition, there is one cation–π interaction between R176 and −36.

Interestingly, the primary base-specific protein–DNA interactions occur at only three positions of the 7-bp −35 element (all Guanines, −35, −34, and −31′) (Figure 3A). The upstream edge of the −35 element is recognized through a series of hydrogen bonds and van der Waals interactions, mostly between R176 and S172 and the guanine bases at −35 and −34. R176 forms two hydrogen bonds with the −35G. In addition, R176 forms a cation–π interaction with the −36 DNA base, creating a hairpin motif along with the −35 hydrogen bonds [24,25]. S172 forms direct hydrogen bond and van der Waals interactions and van der Waals forces (Figures 2 and 3A). In addition, there is one cation–π interaction between R176 and −36.

In contrast to the numerous base-specific interactions at the −35, −34, and −31′ positions, the −33 and −32 positions each contain only one base-specific contact, in the form of van der Waals interactions between the thymidine C5 methyl groups at −33′ and −32′ with F175 and R171, respectively (Figure 3A). The structure reveals no base-specific protein–DNA interactions at the −30 and −29 positions.

### Geometry of the σF4 −35 Element DNA

Over four of the −35 element positions (−33, −32, −30, −29), there are a total of only two protein–DNA base contacts, both weak, van der Waals contacts (Figure 3A). Nevertheless, the −33 and −32 positions are the most highly conserved positions, not only in the Ec σF −35 consensus but also across all group IV σ factors where the promoter specificity is known (Figure 3B; [7,17]). Furthermore, genetic screens for defective transcription resulting from single nucleotide substitutions in the −35 element of the Ec σF homolog from *Salmonella enterica* serovar Typhimurium only resulted in the selection of mutants with substitutions at positions −33 and −32 [26]. Therefore, how is it that the most highly conserved and essential positions in the σF −35 element are also the same ones that lack strong protein–DNA base interactions? The answer for this apparent paradox comes from the unique DNA geometry of the σF −35 element (Figure 4).

The unique DNA geometry induced by oligo(dA)–oligo(dT) tracts, defined by the presence of four to six consecutive A–T bp, is well established [27–31]. Depending on its sequence, oligo(dA)–oligo(dT) tract DNA is rigid and straight, with a high degree of propeller twist and a very narrow minor groove. Despite not being a true oligo(dA)–oligo(dT) tract as a result of the cytosine insertion at −31, the σF −35 element DNA is relatively straight (Figure 4A), with a high degree of propeller twist (Figure S1), and the minor groove width begins to narrow at the start of the −33′–32 AA (Figure 4B). The narrow minor groove is stabilized by a network of cross-strand hydrogen bonds between adjacent DNA bases, along with a spine of hydration consisting of water-mediated hydrogen bonds between the two strands (Figure 4C). The AA at −33′–32 is the most highly conserved feature of the σF −35 consensus. After the −31 cytosine insertion, the consensus comprises TT (−30′–29′). Furthermore, there is a continued run of two additional conserved Ts at −28′–27 (Figure 3B; [17]).

Interestingly, the nucleosome structure [32] contains a stretch of DNA, GAACTT, similar in sequence to −34 to −29 (GAACTT) of the Ec σF −35 element (Figure S2). Similar to Ec σF −35 element DNA, the nucleosome DNA cannot be classified as a typical oligo(dA)–oligo(dT) tracts as a result of the non-A/T base, yet it too displays the hallmark DNA geometry, such as a very narrow minor groove (Figure S2B). The presence of similar DNA geometry in two different structural contexts strongly suggests that the oligo(dA)–oligo(dT)–like DNA geometry found in the Ec σF −35 element DNA complex is an intrinsic property of the DNA sequence and not due to protein induced conformational changes.

The absence of strong, base-specific protein–DNA interactions at the −33, −32, and −30 to −27 positions (Figure 3A) is conspicuous in light of the high DNA sequence conservation, particularly at the −33′–32 positions (Figure 3B). This, combined with the observation that the DNA sequence induces a unique geometry in the −35 element DNA (Figure 4), strongly suggests that the DNA sequence is conserved at these positions to set up the global conformation of the DNA, and that this DNA conformation is essential for σ4 binding.

In this light, the results of the previous genetic screen [26] make good sense. Individual mutations at positions other than the −33 and −32 could be compensated for by both the binding interactions at other −35 element positions and by protein–DNA backbone interactions, which would not be lost at the mutated position. However, substitutions at the −33′–32 positions, which disrupt the highly conserved AA, would in turn disrupt the global DNA geometry necessary for σ4 binding.

### Comparison of σF4 and σA4 −35 Element Recognition

Superposition of the DNA from the Ec σF4 and *Taq* σA4 [4] −35 element complexes reveals that Ec σF4 binds A further into the major groove than the group I σ factor *Taq* σA4, allowing Ec σF4 to form more extensive interactions with the DNA (Figure 5A). In addition, this shift extends the DNA recognition surface of the protein toward the C-terminus of the helix-turn-helix motif recognition helix of Ec σA4 (Figure

---

**Table 2. Ec σF/DNA Crystallographic Analysis and Refinement (against Native Dataset)**

| Space group | P21 |
|-------------|-----|
| Unit cell  | a = 55.009 Å, b = 68.709 Å, c = 61.133 Å, α = 90°, β = 101.254°, γ = 90° |
| Resolution (Å) | 20–23 |
| Number of solvent molecules | 136 H2O |
| R_free/R_refined (%) | 24.07/25.28 |
| RMSD bond lengths | 0.009 Å |
| RMSD bond angles | 1.460° |

R_free = σ2/[F_observed] − [F_calculated]/σ2[F_calculated], R_refined = R_free calculated using 10% random data omitted from the refinement.

DOI: 10.1371/journal.pbio.0040269.t002
5B). For example, even though both promoters have a G at −31, with *Taq* σ^74 in it is recognized by R409 and with *Ec* σ^74 it is recognized by R171, which is four residues (one helical turn) further toward the C-terminus in the aligned sequences.

Furthermore, the aligned residues *Taq* σ^74 K418 and *Ec* σ^74 R176 contact the DNA at different positions. Whereas *Taq* σ^74 K418 makes contacts upstream of the *Taq* σ^74 −35 element at −38, *Ec* σ^74 R176 forms many important interactions within the σ^74 −35 element at −35. Interestingly, *Taq* σ^74 makes one van der Waals and four hydrogen bond protein–DNA contacts upstream of the −35 element at −36 and −38, whereas, *Ec* σ^74 only makes one van der Waals and one cation–π interaction with the nearby −36 DNA base. In essence the 4-Å shift causes the regions of *Taq* σ^74 that were involved in upstream non-promoter element contacts to be involved in sequence specific −35 element contacts in the *Ec*

---

**Figure 2.** *Ec* σ^74/DNA Contacts; Structural View

Two stereo views (front and back) of the *Ec* σ^74 −35 element DNA complex, related by a 180° rotation about the vertical axis as shown. The protein is shown as an α-carbon backbone worm, with σ^74.1 colored yellow and σ^74.2 colored light blue. Side chains are shown for those residues that make protein–DNA contacts. Carbon atoms of the side chains are colored as the backbone, except atoms involved in polar contacts with the DNA are colored (nitrogen atoms, blue; oxygen atoms, red). The DNA is color-coded as in Figure 1A, except atoms involved in polar contacts with the protein are colored (nitrogen atoms, blue; oxygen atoms, red). Water molecules are indicated with red spheres. Dashed black lines indicate hydrogen bonds or salt bridges. DOI: 10.1371/journal.pbio.0040269.g002
Figure 3. Ec σE/DNA Contacts; Schematic View

(A) Schematic representation of σE–DNA interactions for Ec σE (top) and Taq σA (bottom; [4]). The nontemplate/template strand DNA is colored light gray/dark gray (respectively), except the −35 element is colored light green/dark green (for Ec σE) or pink/magenta (for Taq σA). Colored boxes denote protein residues. Color-coding for the proteins, as well as the meaning of the lines indicating interactions, is explained in the legend (lower right). Double thick solid black lines indicate two hydrogen bonds with the same residue. Water molecules mediating protein–DNA contacts are shown as red circles.

(B) Sequence logo denoting sequence conservation within the Ec σE −35 element [17,51].

DOI: 10.1371/journal.pbio.0040269.g003

σF/DNA structure. For example, in both structures aligned residues K418/R176 (Taq σA/Ec σF), T408/P166, R411/T169, and Q414/S172 make up the majority of the upstream nontemplate strand interactions. However, in the case of Ec σF they all make interactions within the −35 element at −35 and −34, whereas in Taq σA they make interactions mostly upstream of the −35 element (−38 to −35). Similarly, the aligned residues R387/R149, L398/Y156, and E399/E157...
interact in both structures with the downstream template strand DNA backbone. However, in EcσE4R149 and E157 make their contacts 1 to 2 bp farther downstream than TaqσA4R387 and E399 (Figure 5B).

In contrast to the genetic screen for nucleotide substitutions in the EcσE/C035 element, which only found decreased transcription from mutations at two of the seven promoter positions (C033 and C032; [26]), systematic mutational studies of the EcσE/C035 element have shown decreased transcription from mutations at five of the six promoter positions (C035 to C031; [33]). The two structures also show major differences in the geometry of the C035 element DNA. Whereas TaqσA bends its C035 element, the protein-bound EcσE/C035 element DNA is relatively straight (Figure 4A). Unlike the C030–C035 element, the EcσE–C035 element itself adopts a unique DNA geometry (described above) that leads to a rigid, straight DNA segment. In fact, unlike the primary Ts, which utilize the flexibility of its C035 element DNA, EcσE appears to use the rigidity of its C035 element DNA sequence to increase specificity.

Superposition of the proteins from the EcσE/C035 element complexes highlights the significant differences in the positioning of the C035 element DNA with respect to the protein, and the different properties of the protein surfaces available for interacting with other proteins bound to the upstream DNA (Figure 5C). Conserved, basic residues of the group I σ domain 4 are key targets for interacting with acidic residues of class II transcriptional activators that bind just upstream of the C035 element [4,34,35]. The role of transcriptional activators in controlling σE transcription is largely unknown.

Implications for C035 Element Recognition by Other Group IV σ Factors

The primary sequences of the group IV σ factors are much more divergent from each other than the members of the other σ70-family subgroups. Furthermore, some genomes contain over 60 group IV σ factors, each of which can recognize unique, but overlapping, sets of promoter sequences. Nevertheless, the various group IV σ factors generally share a high degree of conservation in their C035 element sequences, implying that the less conserved C010 element sequences provide the primary basis for promoter specificity between the different group IV σs, especially within the same species [7,36,37]. Therefore, the mechanism of C035 element recognition revealed in the EcσE/C035DNA structure should be relevant to other group IV σ factors.

Partial to fully characterized regulons have been described for at least eight group IV σs: EcσX [17], Bacillus subtilis (Bsu) σX [38], Bsu σW [39], Pseudomonas aeruginosa (Pae) σX [37,40], Mycobacterium tuberculosis (Mtub) σX [41], Mtub σH [42],...
Streptomyces coelicolor (Scoe) and Pseudomonas syringae (Psyr) HrpL [44]. When considering the −35 elements recognized by these group IV σs together, the −35 element can clearly be divided into three distinct regions. The first is an upstream G region, the second is the previously recognized AAC motif [7], and the third is a less well-conserved downstream T-tract (Figure 6 and Figure S3). The differences and similarities between the consensus −35 elements recognized by these group IV σs can be directly explained from the σ^E_4 sequence alignments in light of the σ^E_4/DNA structure (Figure 6). For example, when consensus sequences for the −35 elements are aligned by the highly conserved AAC motif, all but one of them contain a G at the position equivalent to the Ec −35 position. In the structure, this position is recognized by Ec σ^R_176, which is conserved across all the Group IV σs. At the −34 position of the

**Figure 5. Structural Comparisons of Ec σ^R_4 and Taq σ^A_4 −35 Element Recognition**

(A) Ec σ^R_4/−35 element DNA and Taq σ^A_4/−35 element DNA complexes were aligned using the template strand DNA from −35’ to −30’, giving an RMSD of 0.839 over 30 atoms. The two views are related by a 90° rotation about the horizontal axis as shown. Proteins are shown as α-carbon backbone worms, color-coded as shown. The Ec σ^E_4 −35 element DNA is colored light green (nontemplate strand) and dark green (template strand). The Taq σ^A_4 −35 element is colored pink (nontemplate strand) and magenta (template strand).

(B) Comparison of the Ec σ^E_4 and Taq σ^A_4 protein–DNA interactions. The Ca-backbone of Ec σ^E_4 and Taq σ^A_4 were aligned using Ec σ^E_4 residues 137 to 150 and 155 to 182 with Taq σ^A_4 residues 375 to 388 and 397 to 424, giving an RMSD of 1.00 Å over 42 atoms. Protein residue numbering is shown between the sequences (Taq/Ec). Residues in σ^E_4.1 are highlighted in red/yellow (Taq σ^A_4/Ec σ^E_4) and those in σ^A_4.2 are colored purple/blue. Red dots denote protein residues that make base-specific DNA contacts. Colored dots denote protein residues that make DNA contacts. Black dots denote hydrogen bonds (less than 3.2 Å) or salt bridges (less than 4.0 Å) originating from the protein side chain. Magenta dots denote hydrogen bonds originating from the protein main chain. Blue dots denote van der Waals (hydrophobic) contacts (less than 4.0 Å). Yellow dots denote cation–π interactions. The positions along the DNA that are contacted by each residue are indicated above and below the contact circles.

(C) The protein α-carbon backbones of Ec σ^E_4 and Taq σ^A_4 were aligned as described in (B). The superimposed proteins, shown as α-carbon backbone worms, are shown on the left, color-coded as in (A). The Ec σ^E_4/−35 element and Taq σ^A_4/−35 element complexes are shown separately (middle and left, respectively). In these views, the proteins are shown as molecular surfaces, color-coded according to electrostatic surface potential. The DNAs are shown as phosphate-backbone ribbons, with bases indicated schematically as sticks.

DOI: 10.1371/journal.pbio.0040269.g005

Promoter −35 Element Recognition by σ^E_4
Promoter –35 Element Recognition by $\sigma^F$

The $\sigma^F$–35 element region is highlighted green (similar to $\sigma^E$). All of the Group IV factors have either an F or an H (which could bind oligo(dT)–tract DNA, including oligo(dA)–DNA structure can be applied to a wide range of other group IV $\sigma$ factors. The three –35 element regions are highlighted with the upstream G region (blue), the middle AAC motif (red), and the downstream T rich region (green). Lines connecting the two alignments indicate protein residue–DNA base interactions important for –35 element recognition in the Ec $\sigma^F$/DNA structure.

Figure 6. Correlation of $\sigma_A$ and –35 Element Sequences for Several Group IV $\sigma$ Factors

The top shows a sequence alignment of the proposed –35 element DNA binding region of several group IV $\sigma$ factors. The residue positions that are important in –35 element DNA recognition in the Ec $\sigma^F$–35 element DNA structure are highlighted green (similar to Ec $\sigma^E$) or red (dissimilar to Ec $\sigma^E$). The bottom shows the alignment of the known –35 consensus sequences from several group IV $\sigma$ factors. The three –35 element regions are highlighted with the upstream G region (blue), the middle AAC motif (red), and the downstream T rich region (green). Lines connecting the two alignments indicate protein residue–DNA base interactions important for –35 element recognition in the Ec $\sigma^F$/DNA structure.

DOI: 10.1371/journal.pbio.0040269.g006

Materials and Methods

Cloning, expression, and purification of Ec $\sigma^F_4$. The gene encoding Ec $\sigma^F_4$ (residues 122 to 191) was PCR subcloned from pLC31 [22] into the NdeI/BamHI sites of the pET-15b expression vector (Novagen, Madison, Wisconsin, United States), creating pWJL3. The plasmid was transformed into Ec BL21(DE3)pLysS cells, and transformants were grown at 37 °C in LB medium with ampicillin (100 μg/ml) to an OD$_{600}$ of 0.4 to 0.6. Protein expression was induced with 1 mM IPTG for 4 h.

Cells containing the overexpressed protein were harvested and resuspended in lysis buffer (20 mM Tris-HCl [pH 8.0], 0.5 M NaCl, 5% glycerol, 0.1 mM EDTA, 5 mM imidazole [pH 8.0], 0.5 mM β-ME, and 1 mM phenylmethylsulfonylfluoride). Cells were lysed using a sonicator and clarified by centrifugation. Supernatants were diluted into thrombin digestion buffer (20 mM Tris-HCl [pH 8.0], 0.15 M NaCl, 5% glycerol, 5 mM CaCl$_2$, and 0.5 mM β-ME) and treated with thrombin (500 μg/ml) at 4 °C. To separate the cleaved (untagged) protein from the thrombin and uncleaved, (His)$_6$-tagged protein, the sample was reapplied to the Ni$^2+$-charged HiTrap metal-chelating column (Amersham Biotech [GE Healthcare], Piscataway, New Jersey, United States). Lysis buffer with 20 mM imidazole was used to wash the column, followed by elution of the tagged protein using lysis buffer with 250 mM imidazole. To remove the (His)$_6$-tag, samples were diluted into thrombin digestion buffer (20 mM Tris-HCl [pH 8.0], 0.15 M NaCl, 5% glycerol, 5 mM CaCl$_2$, and 0.5 mM β-ME) and treated with thrombin (500 μg/ml) at 4 °C. To separate the cleaved (untagged) protein from the thrombin and uncleaved, (His)$_6$-tagged protein, the sample was reapplied to the Ni$^2+$-charged HiTrap metal-chelating column in tandem with a 1 ml Benzamidine FF HiTrap column (Amersham), and the flow-through was collected. The sample was

Conclusion

Despite similar function and secondary structure, the group I and IV $\sigma$ factors recognize their –35 elements using distinct mechanisms. The group IV $\sigma$ factor Ec $\sigma^F_4$ binds 4 A further into the major groove than the group I $\sigma$ factor Taq $\sigma^\lambda_4$, making more extensive contacts. Unlike Taq $\sigma^\lambda_4$, Ec $\sigma^F_4$ does not bend the DNA. Instead, conserved sequence elements of the $\sigma^F$–35 promoter induce DNA geometry characteristic of oligo(dA)–oligo(dT)–tract DNA, including pronounced minor groove narrowing. For this reason, the highly conserved AA at –33 to –32 is essential for –35 element recognition by $\sigma^F_4$, even in the absence of direct protein interactions with the DNA bases. It appears that these principles of $\sigma^F$–35 element recognition can be applied to a wide range of other group IV $\sigma$ factors.
then purified using ammonium sulfate (60 g/100 ml sample), centrifuged, and resuspended in gel filtration buffer (20 mM Tris-HCl [pH 8], 0.5 M NaCl, 5% glycerol, and 1 mM DTT). The resuspended sample was applied to a Superdex 75 gel filtration column (Amersham) equilibrated with gel filtration buffer. The eluted Ec σ₄ was concentrated to 30 mg/ml by centrifugal filtration (Vivaspin) and exchanged into a low salt crystallization buffer (20 mM Tris-HCl [pH 8], 0.2 M NaCl, 5% glycerol, 0.1 mM EDTA, and 1 mM DTT). Since Ec σ₄ rapidly precipitated at room temperature when in a low salt buffer (less than 0.5 M NaCl), all subsequent steps were done in the cold room using prechilled supplies. The final purified protein product was aliquoted, flash frozen, and stored at −80 °C. Electrospray mass spectrometry was used to confirm the mass of the purified product (8,427 Da).

### Nucleic acid preparation

For the purposes of crystallization, several DNA constructs were designed, based on the Ec σ₄−35 consensus. Construct length and flanking bases were varied in an attempt to promote crystallization through end-to-end dsDNA contacts. Lyophilized, tritylated, single-stranded oligonucleotides (Oligos Etc., Wilsonville, Oregon, United States) were deprotected and purified on an HPLC using a Varian (Palo Alto, California, United States) Microsorb 300 DNA column [45]. The purified oligonucleotides were dialyzed into 5 mM TEAB (pH 8.5) and dried on a SpeedVac (Savant). The dried oligonucleotides were resuspended in 5 mM Na cacodylate (pH 7.4), 0.5 mM EDTA, 50 mM NaCl to a concentration of 100 nM. Equimolar amounts of oligonucleotides were annealed by heating to 95 °C for 5 min and then cooling to 22 °C at a rate of 0.01 °C/s. The annealed oligonucleotides were dried in a SpeedVac and stored at −20 °C.

### Crystallization and structure determination of the Ec σ₄−35 DNA complex

Co-crystals were obtained by vapor diffusion by mixing the dimer (1:1 ratio) and Ec σ₄ at a final concentration of protein at 1.8 mM (15 mg/ml). The mixture was centrifuged for 30 min, then mixed with an equal volume of well solution (0.4 M MgCl₂, 0.05 M Na-Cacodylate [pH 6.0], and 5% 2-methyl-2,4-pentanediol). Rectangular crystals (0.5 × 0.5 × 0.06 mm) grew within 5 d. Crystals were prepared for cryocryocraphy by soaking in the crystallization solution supplemented with 25% 2-methyl-2,4-pentanediol, followed by flash freezing in liquid nitrogen. A native dataset was collected to 2.3 Å at The National Synchrotron Light Source. The data were processed and scaled using MOSFLM [46] and XDS [47]. The data were further improved by using a 1-hp register offset between the two search model DNAs, to generate a proposed 7-bp DNA which was used to do tandem molecular replacement searches using the 6-bp −35 element from the Taq σ₄−35DNA structure [44; first DNA: Molrep Corr = 0.464 and second DNA: Molrep Corr = 0.475]. In addition to placing the dsDNA density by keeping the Ec σ₄−35 dimer fixed and doing two tandem molecular replacement searches using the 6-bp −35 element from the Taq σ₄−35DNA structure [44; first DNA: Molrep Corr = 0.464 and second DNA: Molrep Corr = 0.475]. In addition to placing the dsDNA into the previously seen DNA density, it extended the density one 4-bp past the DNA search model. The solution was further improved by using a 1-hp register offset between the two search model DNAs, to generate a 7-bp DNA which was used to do two tandem Molrep molecular replacement searches (first DNA: Molrep Corr = 0.464; second DNA: Molrep Corr = 0.487). CNS v1.1 [47] was then used to perform density modification, giving an improved electron density map in which clear density could be seen for the entirety of both dsDNAs, excluding the overhanging base at the downstream end of the DNA. The final DNA was built using a starting template of straight B-form dsDNA corresponding to the crystal structure (constructed using MolmouT2; http://sourceforge.net). Model building was done using O v9.0.7 [48] and refinement using CNS v1.1 (Table 2).

### Accession Numbers

Structure coordinates and structure factors from the Ec σ₄−35DNA crystals have been deposited in the Protein Data Bank (http://www.rcsb.org/pdb) under ID code 2H27. The Protein Data Bank accession number for the nucleosome structure in Figure S2A is 1KX4.

### Acknowledgments

We thank M. Becker and the staff at the National Synchrotron Light Source Beamliner X25 for support, and Tom Muir for access to the electrospray mass spectrometer. We thank Elizabeth A. Campbell, Deepti Jain, Valerie Lannoy, Lars Westblade, Chris Lima, and Tom Muir for helpful discussions and advice.

**Author contributions.** WJL and SAD conceived and designed the experiments. WJL performed the experiments with assistance from SAD. WJL and SAD analyzed the data. WJL wrote the paper, with assistance form SAD.

**Funding.** WJL was supported by National Institutes of Health MSTP grant GM07739 and The W.M. Keck Foundation Medical Scientist Fellowship. This work was supported by National Institutes of Health grant GM53759 to SAD.

**Competing interests.** The authors have declared that no competing interests exist.

---

**Supporting Information**

**Figure S1.** Comparisons of Ec σ₄ and Taq σ₄−35 Element DNA Geometry

(A) Propeller twist, (B) DNA backbone, (C) curvature, and (D) major groove width calculated using 3DNA.

Found at DOI: 10.1371/journal.pbio.0040269.s001 (509 KB TIF).

**Figure S2.** Comparison of Ec σ₄−35 Element DNA and Nucleosome DNA

(A) The nucleosome structure contains a sequence similar to the Ec σ₄−35 Element DNA. Both DNA sequences contain an AA-tract followed by a non-A/T base and then a TT-tract. Despite the non-A/T base, both structures contain narrow minor grooves, which are characteristic of oligo(A)•oligo(T) tracts. The DNA structures were aligned using the template strand phosphates. The minor groove narrowing is evident from the location of the non-template strand DNA relative to B-form DNA. The Ec σ₄−35 element DNA is in green and the nucleosome DNA orange.

(B) Graph showing the DNA minor groove width (calculated using 3DNA) for B-form DNA (blue), Ec σ₄−35 element DNA (green), and nucleosome DNA (orange). Minor groove width was calculated as the P-P distance minus 3.8 Å to take into account the radii of the phosphate groups.

Found at DOI: 10.1371/journal.pbio.0040269.s002 (2.7 MB TIF).

**Figure S3.** Correlation of Ec σ₄ and −35 Element Sequences, along with the −10 Element Consensus, for Several Group IV σ Factors

The top shows a sequence alignment of the proposed −35 element DNA binding region of several Group IV σ factors. The residue positions that are important in −35 element DNA recognition in the Ec σ₄−35 element DNA structure are highlighted green (similar to Ec σ₄) or red (dissimilar to Ec σ₄). The bottom shows the alignment of the known −10 (right) and −35 (left) consensus sequence logos from several Group IV σ factors. The three −35 element regions are highlighted with the upstream G region (blue), the middle AAC motif (red), and the downstream T rich region (green). Lines connecting the two alignments indicate protein residue−DNA base interactions important for −35 element recognition in the Ec σ₄−DNA structure. Despite being more divergent then the −35 elements it is still possible to generate a proposed −10 element alignment. Possible regions of similarity within the −10 elements have been highlighted in light blue, magenta, and gray. The single base change thought responsible for the differential gene regulation between Bsu σ₄ and Bsu σ₄ is indicated with a red arrow. The column to the right of the sequence logos contains the signal and mechanism of regulation for each σ factor.

Found at DOI: 10.1371/journal.pbio.0040269.s003 (1.7 MB TIF).

---

**Protein–DNA contacts were analyzed using the program CON-**

TACT, followed by geometric verification using PyMOL v0.98 (http://www.pymol.org). Cation phosphate groups.

**DNA geometry was analyzed using 3DNA v1.5 [49] and Curves v5.1 (http://www.ibp.fr/UPR9080/Curindex.html). Electrostatic surfaces were calculated using APBS: Adaptive Poisson-Boltzmann Solver [50]. All structural figures were prepared using PyMOL.**
1. Darst SA (2001) Bacterial RNA polymerase. Curr Opin Struct Biol 11: 155–162.
2. Gross CA, Chan C, Dombroski A, Gruber T, Sharp M, et al. (1998) The functional and regulatory roles of sigma factors in transcription. Cold Spring Harb Symp Quant Biol 63: 141–155.
3. Murakami K, Darst SA (2003) Bacterial RNA polymerases: The whole story. Curr Opin Struct Biol 13: 31–39.
4. Campbell EA, Muzzin O, Chlenov M, Sun JL, Olson CA, et al. (2002) Structure of the bacterial RNA polymerase promoter specificity sigma factor. Mol Cell 9: 527–539.
5. Murakami K, Masuda S, Campbell EA, Muzzin O, Darst SA (2002) Structural basis of transcription initiation: An RNA polymerase holoenzyme/DNA complex. Science 296: 1295–1299.
6. Mittenhuber G (2002) A phylogenetic study of the general stress response sigma factor sigmaE of Bacillus subtilis and its regulatory proteins. J Mol Microbiol Biotechnol 4: 427–452.
7. Helmann JD (2002) The extracytoplasmic function (ECF) sigma factors. Adv Microb Physiol 46: 47–110.
8. Lenton M, Gribkov M, Gross CA (1992) The family: Sequence conservation and evolutionary relationships. J Bacteriol 174: 3843–3849.
9. Manganeli R, Provedi R, Rodrigue S, Beaucer J, Gaudreau L, et al. (2004) Sigma factors and global gene regulation in Mycobacterium tuberculosis. J Bacteriol 186: 895–902.
10. Gruber TM, Gross CA (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol 57: 441–466.
11. Raivo TL, Silhavy TJ (2001) Periplasmic stress and ECF sigma factors. Annu Rev Microbiol 55: 591–624.
12. Bartagalone C, Missiakas D, Rains S (2001) Characterization of the Escherichia coli sigmaE regulation. J Biol Chem 276: 20866–20875.
13. De Las Penas A, Connolly L, Gross CA (1997) The sigmaE-mediated response to extracytoplasmic stress in Escherichia coli is transduced by RseA and RseB two negative regulators of sigmaE. Mol Microbiol 24: 573–585.
14. De Las Penas A, Connolly L, Gross CA (1997) sigmaE is an essential sigma factor in Escherichia coli. J Bacteriol 179: 6862–6864.
15. Missiakas D, Mayer MP, Lemaire M, Georgopoulos C, Raina S (1997) Two negative regulators of extracytoplasmic stress in Mycobacterium tuberculosis. J Mol Biol 276: 1053–1061.
16. Alba DM, Gross CA (2004) Regulation of the Escherichia coli sigmaE-dependent envelope stress response. Mol Microbiol 52: 613–619.
17. Rhodius VA, Suh WC, Nonaka G, West J, Gross CA (2000) Conserved and variable functions of the sigmaE stress response in related genomes. PLoS Biol 4: e2.
18. Kovacekova G, Skorupski K (2002) The sigE family: Role and regulation. Mol Microbiol 41: 423–437.
19. Manganeli R, Voskuil MI, Schoolnik GK, Smith I (2001) The Mycobacterium tuberculosis ECF sigma factor sigmaE. Role in global gene expression and survival in macrophages. Mol Microbiol 41: 425–437.
20. Manganeli R, Voskuil MI, Schoolnik GK, Dubnau E, Gomez M, et al. (2002) Role of the extracytoplasmic-function sigma factor sigmaH in Mycobacterium tuberculosis global gene expression. Mol Microbiol 45: 365–374.
21. Paget MS, Molle V, Cohen G, Aharonowitz Y, Buttner MJ (2001) Defining the disulphide stress response in Streptomyces coelicoflavus A3(2): Identification of the sigR regulator. Mol Microbiol 42: 1007–1020.
22. Xiao Y, Hutcheson SW (1994) A single promoter sequence recognized by a newly identified alternate sigma factor directs expression of pathogenicity and host range determinants in Pseudomonas syringae. J Bacteriol 176: 1069–1083.
23. Aggarwal AK (1990) Crystalization of DNA binding proteins with oligo deoxynucleotides. Methods Enzymol 173: 83–90.
24. Vagin A, Teplyakov A (1997) MOLREP: An automated program for molecular replacement. J Appl Crystallogr 30: 1022–1025.
25. Adams PD, Pannu NS, Read RJ, Brünger AT (1997) Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. Proc Natl Acad Sci U S A 94: 5018–5023.
26. Jones TA, Zou JY, Cowan S, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 47: 110–119.
27. Lu XJ, Olson WK (2003) 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nuclear acid structures. Nucleic Acids Res 31: 5108–5121.
28. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. Proc Natl Acad Sci U S A 98: 10077–10082.
29. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. Genome Res 14: 1188–1190.