The bacterial σ factors confer promoter specificity to the RNA polymerase (RNAP). One alternative σ factor, σ^N, is unique in its structure and functional mechanism, forming transcriptionally inactive promoter complexes that require activation by specialized AAA^+ ATPases. We report a 3.4-Å resolution X-ray crystal structure of a σ^N fragment in complex with its cognate promoter DNA, revealing the molecular details of promoter recognition by σ^N. The structure allowed us to build and refine an improved σ^N-holoenzyme model based on previously published 3.8-Å resolution X-ray data. The improved σ^N-holoenzyme model reveals a conserved interdomain interface within σ^N that, when disrupted by mutations, leads to transcription activity without activator intervention (so-called bypass mutants). Thus, the structure and stability of this interdomain interface are crucial for the role of σ^N in blocking transcription activity and in maintaining the activator sensitivity of σ^N.

RNA polymerase | σ^N | σ^N− | transcription | X-ray crystallography

Transcription initiation is a major point for controlling gene expression in all organisms. All cellular RNA polymerases (RNAPs, the central enzyme of transcription) require initiation-specific factors to recognize promoters and generate the transcription bubble. In bacteria, a single RNAP performs all transcription and requires only a single protein, the σ factor, for initiation functions (1, 2). Bacteria have one essential primary or housekeeping σ (σ^70 in Escherichia coli) that controls initiation of the majority of cellular genes, and can have many alternative σ’s that control regulons in response to environmental or physiological cues (3).

All primary σ’s, and almost all alternative σ’s, belong to a homologous protein family, the σ^70 family (4), that functions through a common mechanism. The σ^70-family members bind to the ~400-kDa RNAP catalytic core enzyme (E) to form the σ^70-holoenzyme (En^70). The En^70 then locates promoters that usually contain two hexamer sequence elements, the −10 and −35 elements (normally centered 9.5 and 32.5 nucleotides upstream of the transcription start site, respectively) (1, 2). Upon locating the promoter, En^70 spontaneously isomerizes to the transcription-competent open complex, in which the duplex DNA from the −10 element downstream to the start site (at +1) is unwound to form the transcription bubble, and the DNA template strand (t-strand) is loaded into the RNAP active site (5, 6).

One alternative σ, σ^Aae (also called σ^A ae), is widely distributed among bacteria and is an evolutionary outlier. The σ^Aae has no sequence relationship with the σ^70 family and functions to initiate promoter-specific transcription through a unique mechanism (7). The σ^Aae-holoenzyme (En^Aae) recognizes promoters containing sequence elements distinct from σ^70, the −12 and −24 elements (8) (usually centered −14 and −24.5 nucleotides upstream of the start site; Fig. 1A), and forms a stable complex with promoter DNA that does not spontaneously isomerize to the open complex (9–11). Open complex formation requires activation by eukaryotic-like enhancer-binding proteins that bind far upstream of the promoter, contact the En^Aae closed complex through DNA looping, and catalyze DNA melting in a reaction that requires ATP hydrolysis (12).

Whereas the σ^70-family structure/function relationship has been placed on a firm basis (1, 2, 5, 6), much less is known for σ^Aae.

The overall architecture of En^N was revealed by a 3.8-Å resolution crystal structure (13). The structural basis for recognition of the −24 element by the C-terminal “RpoN” domain (Fig. 1B) was determined through NMR studies (14). Recognition of the −12 element, which is thought to occur through a domain separate from the RpoN domain [the −12-binding domain (−12BD); Fig. 1B], as well as the relationship between the two σ^N promoter-binding determinants, is poorly understood.

Unlike σ^70 factors, σ^N interacts specifically with promoter DNA in the absence of RNAP (15). Here, we determine a 3.4-Å resolution X-ray crystal structure of an Aquifex aeolicus (Aae) σ^N fragment in complex with promoter DNA, revealing the molecular details of σ^N promoter recognition. Moreover, the structure allowed us to build and refine a corrected En^N model using the previously published 3.8-Å resolution X-ray data (13). The corrected En^N structure reveals a conserved interdomain interface within σ^N that harbors all known “bypass” mutants (σ^N substitutions that bypass the requirement for the ATPase activator), providing insight into the maintenance of the closed complex and sensitivity to activator.

Results

**Aae σ^N Forms a Specific Complex with Promoter DNA That Is Stabilized by Truncation of the N-Terminal Region I.** For structural studies, we chose to work with σ^N from the hyperthermophile *Aae*. *Aae* σ^N has previously been purified and characterized; *Aae* σ^N had no sequence relationship with the σ^70 family, and thus, is unique in its structure and functional mechanism, forming transcriptionally inactive promoter complexes with RNAP that require activation by specialized ATPases. The structural basis for σ^N function is of great interest but poorly understood. Here, we determined an X-ray crystal structure of a σ^N fragment bound to promoter DNA, revealing the molecular details of promoter recognition by σ^N. Moreover, the new structure allowed us to build and refine a corrected σ^N-holoenzyme (σ^N/RNAP) complex model using previously published X-ray data. This work overall provides a solid structural framework with which to address further the poorly understood mechanism of activator function in ATP hydrolysis-dependent promoter opening.

Significance

The bacterial σ factors confer promoter specificity to the RNA polymerase (RNAP). One σ factor, σ^N, is unique in its structure and functional mechanism, forming transcriptionally inactive promoter complexes with RNAP that require activation by specialized ATPases. The structural basis for σ^N function is of great interest but poorly understood. Here, we determined an X-ray crystal structure of a σ^N fragment bound to promoter DNA, revealing the molecular details of promoter recognition by σ^N. Moreover, the new structure allowed us to build and refine a corrected σ^N-holoenzyme (σ^N/RNAP) complex model using previously published X-ray data. This work overall provides a solid structural framework with which to address further the poorly understood mechanism of activator function in ATP hydrolysis-dependent promoter opening.

Author contributions: E.A.C. and S.A.D. designed research; E.A.C., S.K., M.W., and S.A.D. performed research; E.A.C., K.R.R., and S.A.D. analyzed data; and E.A.C. and S.A.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The X-ray crystallographic coordinates and structure factor file for the *Aquifex aeolicus* σ^N/RNA polymerase DNA structure have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 5J1B). The coordinates for the corrected En^N model have also been deposited in the Protein Data Bank (PDB ID code 5J1B).
σ\textsuperscript{N} supports promoter-specific transcription in combination with E. coli (Eco) RNAP and an Aae activator in vitro (16). Aae σ\textsuperscript{N} lacks the nonessential region II (RII) (Fig. 1B), which may comprise extended loops recalcitrant to crystallization in the absence of the RNAP (13). High-resolution NMR structures of the core-binding domain (CBD; Fig. 1B), the RpoN domain, and the RpoN

Fig. 1. Architecture of σ\textsuperscript{N}, its promoter, and the effects of −12 element base substitutions. (A) Sequence logo (60) generated from a compilation of 85 σ\textsuperscript{N} promoters with experimentally confirmed start sites (8). The sequence of the Aae dhsU promoter (16) is shown below the logo. The −24 and −12 elements are shaded yellow. (B) Schematic diagram showing σ\textsuperscript{N} domain architecture. The sequence numbering (below the diagram) corresponds to Aae σ\textsuperscript{N} numbering (which lacks RII). The ΔNσ\textsuperscript{N} is also denoted. (C) Plot showing the effects (in kT units) of −12 element base substitutions on the binding free energy (per molecule) relative to the wild-type (wt) promoter. Shown above is the sequence of the dhsU promoter, with the −24 and −12 elements in red font. The individual −12 element base substitutions are denoted (SI Appendix, Fig. S1A). anti, −12 antisequence. Data for each mutant represent the average of at least three independent measurements, and the error bars represent SEM.
domain in complex with -24 element DNA were determined using Aae σN domains (14, 17, 18). The DNA templates for most of our biochemical and structural studies were based on the Aae dhsU promoter (16), which corresponds to the consensus σN promoter (Fig. 1A). Both purified Aae σN and Aae σD with the N-terminal RI-truncated (ΔNeN; Fig. 1B) form specific complexes with a 36-nt dhsU template (SI Appendix, Fig. S1). Full-length Aae σN forms a complex with a dissociation constant (Kd) of 460 nM (Table 1), as determined by quantitative filter-binding assay. As expected (19, 20), Aae ΔNeN bound more tightly, forming a specific complex with a Kd of 100 nM (Table 1).

The -12 Element Plays an Important Role in the Aae σN/Promoter Complex. Previous studies suggested that σN RpoN domain/-24 element interaction plays a much more important role than interactions with the -12 element in providing binding stability for the σN/promoter DNA interaction (21–23). The high sequence conservation of the -12 element (Fig. 1A) points to its importance, but evolution does not necessarily select promoter sequences based on binding affinity alone. To test the importance of the -12 element sequence for the stability of Aae σN/DNA complexes, we generated a -12 anticonsensus promoter template comprising the dhsU sequence with the exception that conserved bases within and near the -12 element were substituted with the least likely bases to occur in that position (SI Appendix, Fig. S1A). Substitution of the -12 element with the anticonsensus -12 element significantly destabilized the complexes with Aae σN and ΔNeN (-20-fold increase in Kd corresponding to ~3 kT units, where k is the Boltzmann constant and T is temperature; Fig. 1C and Table 1).

We also tested the importance of individual -12 element bases by substituting them one at a time with the least likely base to occur at each position and measuring the binding affinity with σN and ΔNeN (Fig. 1C and Table 1). The substitution −16TA had very little influence on the binding energetics, whereas the substitutions −15TA, −14GA, and −13CG had moderate effects (mostly between 1.5 and 2 kT units, and similar for both proteins). The −12AC substitution had a moderate effect for σN, but less for ΔNeN, suggesting that σN RI plays a role in recognition of the most downstream edge of the -12 element. Overall, we conclude that favorable contacts between conserved bases of the -12 element and residues of σN contribute significantly to the stability of σN and ΔNeN complexes with promoter DNA.

The -12 Element Is Recognized by a Winged Helix–Turn–Helix DNA-Binding Domain. To provide insight into the recognition of the -12 element, as well as the relationship between -24 and -12 binding determinants of σN, we determined the crystal structure of Aae ΔNeN in complex with a DNA fragment containing the Aae dhsU promoter sequence to 3.4-Å resolution (Fig. 2 and SI Appendix, Fig. S2 and Table S1). The crystals contain two protein/DNA complexes per asymmetric unit with a root mean square deviation (rmsd) of 0.95 Å over 286 α-carbons between them. The ΔNeN comprises three structured domains, the CBD, the −12BD, and the RpoN domain, linked by flexible linkers (Figs. 1B and 2B).

Within the −12BD, the −12 element is recognized by the helix–turn–helix (HTH) DNA-binding motif of a winged-HTH domain elaborated with an N-terminal β-hairpin, a loop, and a long α-helix [termed extra-long helix (ELH) by Yang et al. (13)], and with a C-terminal β-hairpin that forms a four-stranded β-sheet with the N-terminal β-hairpin (SI Appendix, Fig. S3A). The recognition helix of the HTH motif inserts into the major groove of the −12 element and establishes sequence-specific contacts (Figs. 2B and S3B).

The −12BD is anchored to the DNA duplex through phosphate backbone interactions with the nontemplate (nt) strand (Fig. 3A). The positively charged side chain of R295 makes polar interactions with the −18 nt-strand [−18(nt)] and −17(nt) phosphates. The peptide backbone amides of L294 and R295, at the beginning of HTH-motif helix 1, form a hydrogen bond with the −17(nt) phosphate, an interaction typical for HTH motifs (24). Conserved S309 (Fig. 3D) interacts with the −16(nt) phosphate, contributing a modest 0.7 kT units to binding energy (Table 2). In addition, R327, which is not within the HTH motif but is located just after the C-terminal β-hairpin of the wing, interacts with the −16(nt) phosphate. This interaction is likely important because R327 is conserved as either K or R (Fig. 3D).

The overall sequence of the −12BD is not well conserved, but the conservation of a seven-residue sequence that includes part of the HTH-motif recognition helix (residues 305–310, −HESTYSR–) is striking, suggesting functional importance (Fig. 3D). We substituted the residues within this segment that had the potential to contact DNA (H304, E305, S306, T307, S309, and R310) within the context of the Aae ΔNeN truncation and tested promoter binding. Substitutions of H304, S306, T307, and S309 with Ala had little (H304 and S306) to only mild (T307 and S309) effects on promoter binding (Table 2). By contrast, substitutions of E305 and R310 had strong but opposite effects.

Substitution of E305 by several other residues (E305A, E305Q, or E305S) resulted in significantly tighter promoter binding, decreasing the Kd by more than an order of magnitude (~2.7 to ~3.7 kT units; Table 2). In the structure, E305 does not directly contact the DNA, but the negatively charged residue is inserted deep within the major groove (Fig. 3B and C).

R310 is absolutely conserved (Fig. 3D), and the substitution R310A disrupts promoter binding by ΔNeN, increasing the Kd by more than an order of magnitude (2.6 kT units). In the structure, R310 makes base-specific hydrogen bonds with G14(nt) (Fig. 3A–C) of the nearly universally −12 element GC motif.

Table 1. Aae σN and ΔNeN binding to dhsU promoter mutants

<table>
<thead>
<tr>
<th>Sigma</th>
<th>Aae σN/Kd</th>
<th>Full-length Aae σN/Kd</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Kd nM</td>
<td>kT units [ln(Kd/ Kd wt)]</td>
</tr>
<tr>
<td>Consensus*</td>
<td>103 ± 15</td>
<td>103 ± 15</td>
</tr>
<tr>
<td>-12 anticonsensus*</td>
<td>1,900 ± 270</td>
<td>2.9 ± 0.2</td>
</tr>
<tr>
<td>-12AC</td>
<td>167 ± 52</td>
<td>0.49 ± 0.34</td>
</tr>
<tr>
<td>-13CG</td>
<td>498 ± 49</td>
<td>1.6 ± 0.2</td>
</tr>
<tr>
<td>-14GA</td>
<td>322 ± 19</td>
<td>1.1 ± 0.2</td>
</tr>
<tr>
<td>-15TA</td>
<td>641 ± 71</td>
<td>1.8 ± 0.2</td>
</tr>
<tr>
<td>-16TA</td>
<td>110 ± 24</td>
<td>0.07 ± 0.3</td>
</tr>
</tbody>
</table>

mut, mutated; wt, wild type.

*SI Appendix, Fig. S1A.

1Kd value (in nM) ± SEM.
and

RpoN domain alone binds preferentially to so-called –Aae and promoter DNA, and 2 promoter DNA complex. Aae dhsU or E structure σ (i.e., –σ or –σ).

Crystal structure of the σB σσ−σ or −σ. The interaction of full-length RpoN with promoter DNA is shown in CPK format and color-coded as in Fig. 2. Overall structure of the σB σσ−σ or −σ (Fig. 1). The structural and mutagenesis results combined indicate –12BD matches −12 element (14) (Fig. 2). The interaction of full-length RpoN with promoter DNA is shown in CPK format and color-coded as in Fig. 2. Overall structure of the σB σσ−σ or −σ (Fig. 1). The structural and mutagenesis results combined indicate −12BD makes only a significant distortion by Er binding. This conclusion is strengthened by the finding that ErN or oN alone binds preferentially to so-called “early-melted intermediate” promoter templates, containing an engineered bubble through nt-strand noncomplementarity from −12 to −11 (29, 30). ErN or ΔErN does not induce DNA reactivity with chemical probes and binds preferentially to fully duplex over early-melted intermediate promoter DNA, indicating that the DNA distortion downstream of the −12 element GC motif can be attributed to RI.

In our superposition of RI onto the ΔErN/dhsU DNA complex, RI sterically clashes with the duplex promoter DNA from −12 to −9 (Fig. 4B). We hypothesize that in the binary complex with σN or ErN, the base pairs of the promoter DNA from −12 to −9 are distorted due to the presence of RI, likely kinking the downstream DNA.

Corrected ErN Model. Structures of each of the ΔErN domains were determined previously in different contexts:

1) CBD: The isolated Aae σN-CBD was determined by solution NMR [Protein Data Bank (PDB) ID code 2KM9] (18). The Klebsiella pneumoniae (Kpn) σN-CBD was determined in the context of the 3.8-Å resolution X-ray crystal structure of ErN (PDB ID code 5BYH) (13). As noted by Hong et al. (18), the CBD comprises two subdomains (CBDa and CBDb; Figs. 1B and 2B) that can flex with respect to each other, so structural comparisons were done within each subdomain. Our structure of the Aae CBD matches PDB 2KM9 (maximum rmsd of 1.5 Å over 64 aligned Ca’s for CBDa; SI Appendix, Table S2), but does not match PDB 5BYH (rmsd of 2.4 Å over 49 Ca’s for CBDa and 5.8 Å over 36 Ca’s for CBDb; SI Appendix, Table S2).

Fig. 2. Crystal structure of the Aae ΔErN/dhsU promoter DNA complex. (A) Oligonucleotides used for crystallization, derived from the Aae dhsU promoter sequence (16). (B) Overall structure of the Aae ΔErN/DNA complex. The protein is shown as a backbone ribbon and color-coded as in Fig. 18. The DNA is shown in CPK format and color-coded as in A. Segments of the protein structure connected by disordered loops are schematically illustrated with colored balls (CBD to −12BD linker, residues 192–194, gray; −12BD, residues 239–241, green; RpoN domain, residues 353–356, brown). Two orthogonal views are shown. (Upper) CBD and CBD (−12BD) linker are removed for clarity.

The −24 Element Is Recognized by the HTH Motif of the RpoN Domain. Linked to the −12BD by a five-residue, flexible linker is the RpoN domain, a canonical HTH motif that recognizes the −24 element (14) (Fig. 2B). The RpoN box (residues 377–386; −ARRTVAKYRE–), contained within the recognition helix of the RpoN-domain HTH motif, is the most conserved amino acid sequence element in σN (25). The protein/DNA interactions observed in the crystal structure are very similar to the interactions seen in an NMR structure of this same Aae RpoN domain bound to a −24-element–containing DNA oligonucleotide (14). These interactions include interactions of T380, Y384, and R385 with the DNA phosphate backbone, and sequence-specific read-out of the essentially absolutely conserved −24-element GG motif (Fig. 1A) by R378 [G.26(nt)] and R379 [G.26(nt)] (Fig. 3A and B).
Fig. 3. Protein/DNA interactions in the Aae ΔNσN/DNA structure. (A) Schematic representation of protein/DNA interactions. The DNA is color-coded as in Fig. 2A. The orange rectangles denote protein residues contacting the DNA. Colored lines denote interactions: yellow, van der Waals (≤4.5 Å); green, H-bond (≤3.5 Å); red, salt bridge (≤4.5 Å). A water molecule mediating an interaction between S309 and the −15(nt) phosphate backbone is shown as a pink sphere. (B) Overall view showing the RpoN domain/−24 element and −12BD/−12 element interactions. Polar interactions (H-bond or salt bridge) are denoted by dashed gray lines. Selected residues and bases discussed in the main text are labeled. (C) Close-up view of −12BD/−12 element interactions. Polar interactions (H-bond or salt bridge) are denoted by dashed gray lines. (D) Sequence logo (60) generated from an alignment of 217 σN orthologs. The secondary structure of Aae ΔNσN is denoted above (tube, α-helix; arrow, β-strand). The two α-helices comprise the −12BD HTH motif. Green dots denote residues that interact with the DNA. Two individual σN sequences (Aae σN and Kpn σN) are shown below. The numbering in the context of each sequence is shown at the bottom (Aae σN/Kpn σN).
Table 2. Binding of Aae ΔNeN mutants to consensus dhsU promoter

<table>
<thead>
<tr>
<th>Aae ΔNeN</th>
<th>K_d nM</th>
<th>kT units [ln(K_d^mut/K_d^wt)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wt</td>
<td>103 ± 15*</td>
<td></td>
</tr>
<tr>
<td>H304A</td>
<td>81.1 ± 7.9</td>
<td>-0.24 ± 0.17</td>
</tr>
<tr>
<td>E305A</td>
<td>2.5 ± 1.2</td>
<td>-3.7 ± 0.5</td>
</tr>
<tr>
<td>E305Q</td>
<td>6.7 ± 1.7</td>
<td>-2.7 ± 0.3</td>
</tr>
<tr>
<td>E305S</td>
<td>3.7 ± 0.4</td>
<td>-3.2 ± 0.2</td>
</tr>
<tr>
<td>S306A</td>
<td>111 ± 18</td>
<td>0.08 ± 0.21</td>
</tr>
<tr>
<td>T307A</td>
<td>275 ± 16</td>
<td>0.99 ± 0.15</td>
</tr>
<tr>
<td>S309A</td>
<td>204 ± 43</td>
<td>0.69 ± 0.21</td>
</tr>
<tr>
<td>R310A</td>
<td>1,420 ± 250</td>
<td>2.6 ± 0.2</td>
</tr>
</tbody>
</table>

*K_d value (in nM) ± SEM.

ii) –12BD: The Kpn σ^N(−12BD) was determined in the context of the 3.8-Å resolution X-ray crystal structure of EoN (PDB 5BYH) (13). Our structure of the Aae−12BD does not match 5BYH (rmsd of 8.4 Å over 103 Cα’s; SI Appendix, Table S2).

iii) RpoN domain: The Aae σ^N-RpoN domain was determined by solution NMR in isolation (PDB ID code 2AHQ) (17) and bound to −24 element DNA (PDB ID codes 2O8K and 2O9L) (14). The Kpn σ^N-RpoN domain was determined in the context of the 3.8-Å resolution X-ray crystal structure of EoN (PDB 5BYH) (13). Our structure of the Aae RpoN domain matches all of these structures (SI Appendix, Table S2).

Overall, our structures of the Aae σ^N domains are consistent with the available solution NMR structures of the Aae CBD and RpoN domains, and with the Kpn RpoN domain from PDB 5BYH, but they are not consistent with the Kpn CBD and −12BD from PDB 5BYH (SI Appendix, Table S2). More detailed examination of the Aae σ^N domains and the corresponding Kpn σ^N domains from PDB 5BYH revealed potentially large registration errors in segments of the Kpn CBD, and large registration errors, as well as topology errors, in the Kpn −12BD (Figs. S3 and S4). Although the Aae and Kpn σ^N sequences have diverged significantly (the overall sequence identity is 26%, compared with 90% between Kpn and Eco σ^N), both Aae and Kpn σ^N belong to a large protein family and function in combination with Eco RNAP (16), so such significant structural differences between the Aae and Kpn σ^N domains are unexpected. Using the deposited structure factors for PDB 5BYH, we completely rebuilt and re-refined the EoN structure in light of our Aae ΔNeN structure (details are provided in Materials and Methods). Our EoN model is significantly improved based on three criteria:

i) Crystallographic statistics: We compared the refinement statistics for our corrected EoN model against three sets of refinement statistics associated with PDB 5BYH: the statistics reported in table S1 of ref. 13, the statistics reported in the header of the PDB 5BYH file (which differed slightly), and statistics from a re-refinement of PDB 5BYH (without rebuilding) using PHENIX (31), with exactly the same refinement parameters as used for the refinement of our corrected EoN model. In each case, the statistics for the corrected EoN model are significantly improved for every parameter except the clashscore (SI Appendix, Table S3); the low clashscores for the refinements associated with the original PDB 5BYH are explained by the fact that a huge number of the modeled EoN residues do not include sidechain atoms (so are modeled as Ala). Improvements were seen in the statistical parameters measuring both the correspondence of the corrected EoN with the X-ray data (R_work/R_free decreased by more than 0.0673/0.0272, and CC_work/CC_free increased at least 0.107/0.115 for the corrected EoN model compared with the other models) as well as the quality of the model geometry (SI Appendix, Table S3).

ii) Sequence alignment criteria: Structure-based sequence alignments were generated from the structural superpositions (SI Appendix, Table S2) of the corresponding domains from the Aae σ^N/DNA structure (Fig. 2B), Kpn σ^N from PDB 5BYH, and Kpn σ^N from the corrected EoN model. Using Aae σ^N as a reference, the sequence alignments with the PDB 5BYH Kpn σ^N domains yielded 16.9%, 17.5%, and 43.3% sequence identity for the CBD, −12BD, and RpoN domains, respectively. The same sequence alignments with the corrected Kpn σ^N model yielded 23.8%, 27.7%, and 43.3% sequence identity for the CBD, −12BD, and RpoN domains, respectively.

Fig. 4. Correspondence of the Aae ΔNeN/DNA structure with previous footprinting data. (A) Crystallized DNA sequence (as in Fig. 2A), with previous footprinting results denoted schematically. Dots above (nt-strand) or below (t-strand) denote backbone positions protected from hydroxyl-radical cleavage by σ^N (weak hydroxyl-radical protection, cyan; strong hydroxyl-radical protection, blue) (28). Bases colored red denote guanosines protected from DMS modification by EoN (27). Positions with increased reactivity to KMnO_4, DEP, and/or ortho-copper phenanthroline in the presence of σ^N or EoN are colored orange (20). These base pairs are thought to be distorted or melted in the early melted intermediate (29, 30). (B) Footprinting results in A are mapped onto the DNA in the context of the Aae ΔNeN/DNA complex. Segments of the protein structure connected by disordered loops are schematically illustrated with colored balls (−12BD, residues 239–241, green; RpoN domain, residues 353–356, brown). The −24 and −12 elements are outlined in yellow.
43.3% sequence identity. Clearly, the corrected EσN model better reflects the sequence relationship between the Aae and Kpn σN CBD and −12BD domains.

iii) Fourier difference maps: We calculated two sets of Fourier difference maps:

\[ 2F_o - F_c + σc \] and \[ F_o - F_c \], where \( F_o \) is the observed amplitude (downloaded from the PDB), \( F_c + σc \) are amplitudes and phases calculated from PDB 5BYH, and \( F_c \) are amplitudes and phases calculated from our corrected EσN model. The difference maps calculated using the PDB 5BYH coordinates are inconsistent with the coordinates themselves, whereas the difference maps calculated using the corrected coordinates are consistent with the corrected coordinates (SI Appendix, Figs. S4).

Rebuilding the EσN model required minor adjustments to core RNAP, but major adjustments were required in σN, including some regions with registration errors as large as five positions in the sequence (SI Appendix, Figs. S5A) and topology errors resulting in some α-carbon positions being more than 15 Å out of place (SI Appendix, Fig. S5B).

Specifically, by domain:

i) RI of PDB 5BYH from residues 26–55 had an incorrect sequence register.

ii) The CBD of PDB 5BYH was incorrectly built (mostly sequence register errors) between residues 156–168 and 185–230.

iii) The −12BD of PDB 5BYH contained topology errors and large register shift errors. In PDB 5BYH, the incoming linker from the CBD was connected to the (misassigned) first β-strand of a four-strand antiparallel β-sheet that preceded the first −12BD α-helix (SI Appendix, Fig. S2B).

The overall sequence conservation of the −12BD HTH (Kpn σN residues 315–356) is not particularly striking, with the exception of one position that is absolutely conserved, Kpn σN R336 and K388 (Aae σN R264 and K315) (36, 37). Neither of these residues plays a role in DNA binding in the Aae EσN/DNA structure (Fig. 3). Examination of the corrected EσN model provides insight into the role these residues play in σN function. For the discussion below, we use Kpn σN numbering.

The overall sequence conservation of the −12BD ELH (Kpn σN residues 315–356/Aae σN residues 243–277) is not particularly striking, with the exception of one position that is absolutely conserved, Kpn σN R336 and K388 (Aae σN R264 and K315), the other −12BD bypass mutant position, is located just after the recognition helix of the −12BD HTH motif and is also absolutely conserved (Fig. 3D). These two residues take part in a network of interactions with other conserved residues that appear to stabilize the interaction of RI with the −12BD (Fig. 5). Kpn σN R336 forms a buried salt bridge with D275 (absolutely conserved as a negatively charged residue; Fig. 5C), whereas K388 forms a hydrogen bond with the carbonyl oxygen of R1-L26 (Fig. 5B).

The neighboring position in RI, L25, is conserved as a hydrophobic residue (Fig. 5C), and substitutions at this position give rise to the bypass phenotype (32, 35). Conserved hydrophobic residues from RI (Kpn σN L25, L26, L33, and L37) and the −12BD (Kpn σN L297, L329, and L333, as well as the alkyl chain of R336) all form a complementary hydrophobic interface between RI and the −12BD (Fig. 5).

Discussion

The structural basis for σN function is of great interest because of the distinct initiation mechanism compared with the better understood σ70 family (7, 12). Structural information on σN was limited to NMR structures of isolated domains [CBD (18) and RpoN domain (14, 17)] and a 3.8-Å resolution crystal structure of EsN that revealed the overall disposition of the σN domains in complex with RNAP but is shown here to contain significant errors in detail (Figs. S3–S5 and SI Appendix, Tables S2 and S3).

The details of −24 element recognition by the RpoN domain were delineated by an NMR structure (14), but −12 element recognition had not previously been visualized (Fig. 3).

−12 Element Recognition. In addition to the HTH DNA-binding motif of the RpoN domain, an HTH motif within the −12BD was predicted from early sequence analysis of σN homologs (41), and the −12BD HTH was subsequently structurally confirmed by Yang et al. (13). Site-directed mutagenesis identified residues within the −12BD HTH important for promoter recognition (42), and allelespecific suppression studies identified the −12BD HTH as recognizing the promoter −12 element (26). Our 3.4-Å resolution Aae ΔσN/DNA structure confirms these conclusions (Fig. 3). In fact, numerous details of the early site-directed mutagenesis results (26, 42) are explained by the structure:

i) E305: Multiple substitutions of the residue corresponding to Aae σN E305 were compatible with basic σN function, leading this position to be labeled “promiscuous” (42). In our analysis, substitution of E305 with A, Q, or S dramatically increased

Actuator Bypass Mutants. The requirement for an extrinsic activator and ATP hydrolysis makes the mechanism of σN-mediated transcription initiation strikingly different from the predominant σ70-family mechanism (1, 7). Mutants of σN that allow open complex formation and transcription initiation without activator intervention (so-called bypass mutants) have been described (32–37). RI of σN has been proposed to contact activator directly (38, 39), and plays a critical role in mediating the response of the EsN promoter complex to activator, and most bypass mutants comprise deletions or amino acid substitutions within RI (32, 34, 35, 40).
DNA-binding affinity (Table 2). In the Aae $\Delta N/\sigma N$ DNA structure, E305 does not interact with the DNA but inserts deep into the major groove, explaining how removal of the negative charge resulted in increased DNA-binding affinity. Glu is conserved at this position (Fig. 3), but multiple substitutions at this position increase DNA-binding affinity (Table 2) and apparently do not interfere with basic activator-mediated initiation (42). We suggest that Glu at this position may be important for precise regulation/tuning of $\sigma N$ activity in vivo.

ii) S306: S306 can only be substituted by Ala or Thr (must be a small side chain) (42). S306 sits deep in the major groove and closely approaches $-15T(nt)$ and $-14C(t)$ (Fig. 3), leaving no room for larger residues.

iii) T307: T307 can only be substituted by Ser (42). T307 sits deep in the major groove and makes van der Waals contacts with $-12T(t)$ (Fig. 3), so larger residues cannot be accommodated at this position.

iv) R310: R310 is absolutely conserved (Fig. 3D) and is required for promoter binding (Table 2) and $\sigma N$ function (42). Allele-specific suppression genetics suggested this residue specifically recognized the nearly absolutely conserved $-14GC$ base pair of the $-12$ element (26) (Fig. 4). Our structure shows that R310 makes a base-specific contact with $-14G(nt)$ (Fig. 3).

Overall, there is thus excellent agreement between the previous site-directed mutagenesis results (26, 42), our limited mutagenesis results (Table 2), and the structure (Fig. 3).

Interactions between promoter DNA and $\sigma N$ or En$N$ have been extensively monitored by footprinting approaches probing protection of the DNA from cleavage (Dnase I or hydroxyl-radical) or...
chemical (DMS) agents, or induction of DNA distortions (KMnO₄, DEP, ortho-copper-phenanthroline). The footprinting results from the −30 position to −14 position of the promoter DNA are explained by the Aae ΔNeN/DNA structure, and a model of full-length σ⁰, generated by superimposing σ⁰-RI based on the ErN/DNA structure, fully explains the footprinting results from −30 to −5 (Fig. 4). The correspondence of our Aae ΔNeN/DNA structure to previous mutagenesis and footprinting analyses indicates that the structure recapitulates the solution σ⁰/DNA interactions for both σ⁰ and ErN.

Contrary to expectations, we did not observe any base-specific interactions with the nearly absolutely conserved −13CG bp (Figs. 1A and 3). This lack of base-specific interactions is surprising, because substituting the −13C with G (the least likely base to occur at that position of σN promoters) (8) results in a significant binding defect for both σN and ΔNeN (Fig. 1C). This result is not explained by specific interactions observed in our structure and may indicate that altering this base pair may affect neighboring base pairs or the overall conformation of the DNA in ways that are deleterious to binding. Moreover, this base pair might play an important role in steps subsequent to initial promoter binding that are important for DNA melting. Extensive mutagenesis of the −12 promoter bases has shown that the −13CG bp is critical for the bypass activity as well as regulation (43), suggesting that conservation in the promoter sequence is critical for functions other than initial binding.

Corrected ErN Model. Comparison of the Aae ΔNeN structure (Fig. 2) with the corresponding domains of Kpn σ⁰ from the ErN structure (PDB 5BYH) (13) pointed to the possibility of significant errors in the Kpn σ⁰ model of PDB 5BYH (SI Appendix, Table S2), and, indeed, complete rebuilding and re-refinement of the structure against the deposited X-ray data yielded a corrected ErN model more consistent with the X-ray data (SI Appendix, Fig. S4 and Table S3) and, at the same time, with improved geometry (SI Appendix, Table S3). The higher resolution Aae ΔNeN structure guided the remodeling of the corresponding domains of Kpn σ⁰ (CBD and −12BD). The resulting (improved) models yielded improved Fourier difference maps that allowed us to remodel Kpn σN-RI as well (even though RI was not present in the Aae ΔNeN structure).

Bypass Mutations. The corrected ErN model allowed the identification of a conserved, mostly hydrophobic RI/−12BD interface (Fig. 5 B and C) and provided insight into activator bypass mutants in both RI and the −12BD. ErN binds promoter DNA and generates the early-melted intermediate, but further promoter melting and/or loading of the DNA into the RNAP active site cleft is blocked by the combined presence of RI and the −12BD (13). Interaction with activator and ATP hydrolysis is required to relieve this inhibition of open complex formation. All of the activator bypass mutations disrupt conserved elements of the RI/−12BD interface, indicating that the structure and stability of this interface are crucial for the role of RI in blocking open complex formation until the action of the ATP-dependent activator.

Structure and Function of σ Factors. Comparison of the Aae ΔNeN structure (in the context of the DNA complex; Fig. 2) with the corresponding domains of the Kpn σ⁰ structure (in the context of ErN; Fig. 5A) revealed substantial domain reorientation (SI Appendix, Fig. S6) facilitated by flexible linkers connecting the structural domains (Fig. 1B). This architecture is remarkably similar to that of the σ⁰ family of σ factors (2): The σ⁰ family members comprise a series of structured domains linked by flexible linkers (44, 45), and large rearrangements of the domains with respect to each other occur during aspects of their function (46–48). The σ⁰ and σ⁰ families are not evolutionarily related, so these parallels suggest that large domain rearrangements, and the structural architecture that facilitates them, are irreducible aspects of bacterial σ-factor function.

The primary (group I) σ factors, such as Eco σ⁰, spontaneously nucleate DNA melting (1, 49, 50), necessitating mechanisms to autoinhibit DNA binding in the absence of the core RNAP (51–54). These mechanisms rely on large domain rearrangements upon core RNAP binding to relieve the autoinhibition of DNA binding (46, 47) and to present promoter-recognition elements within different σ domains with the proper orientation and spacing (2, 55). These considerations do not seem to apply to σ⁰, because σ⁰ can bind its own promoters in the absence of RNAP (15) (Fig. 2 and SI Appendix, Fig. S1).

Both σ⁰ and σN have very high binding affinities to core RNAP, with Kd values estimated to be subnanomolar (56). Nevertheless, the σ factors must dissociate from core RNAP upon the transition from initiation to elongation. In the case of σ⁰, the individual structural domains do not interact tightly with core RNAP and the simultaneous interaction of the linked domains gives rise to the high affinity, because all of the domains would have to release simultaneously in order for σ⁰ to dissociate. However, the structural architecture allows for the staged disruption of a series of weak interactions by the progressive elongation of the DNA transcript (2, 45). This mechanism is likely to operate in the case of σ⁰; As with σN, the individual σ⁰ domains do not interact strongly with RNAP and structural elements of σ⁰ occupy the active site channel (RII) (13) and the RNA exit channel (RII-CBD linker and the CBD itself) (13).

Finally, once engaged with promoter DNA, Eco σ⁰ spontaneously isomerizes to the transcription-ready open complex (1). By contrast, the initial ErN/promoter complex requires remodeling by the ATP hydrolysis-dependent action of the activator (7, 12), a process likely facilitated by the flexible domain organization of σ⁰. Our structure and associated analysis of the Aae ΔNeN/DNA complex provided insight into σ⁰ promoter recognition and allowed us to generate a much improved ErN model from previously published data (13). Overall, this work provides a solid structural framework to address the poorly understood mechanism of activator function in ATP hydrolysis-dependent promoter opening further.

Materials and Methods

Full details of the methods used are presented in SI Appendix, SI Materials and Methods.

Cloning, Protein Expression, and Purification. Cloning, protein expression, and purification used standard methods.

Radioactive Filter-Binding Assay. Radiolabeled dsDNA promoter DNA or mutant derivatives (Tables 1 and 2) were mixed in reactions (40 μL) with 0.625 nM DNA and varying concentrations of full-length σN, ΔNeN, or ΔNeN mutants (Table 2). Sterilized 0.45-μm nitrocellulose filters (Millipore) were placed in a vacuum filter-binding assay apparatus (Millipore) and prewashed, and the reaction mixtures (35 μL) were then pipetted onto the filter and washed with 5 mL of buffer. Filters were air-dried and visualized by phosphor imagery. Filter-bound complexes were quantified using ImageQuant 5.2.

Crystalization of the Aae ΔNeN/DNA Complex. Promoter DNA was added to ΔNeN (1.2:1 molar ratio), the complex was concentrated to a final protein concentration of 10 mg/mL by centrifugal filtration, and crystals were obtained by sitting drop vapor diffusion at 22 °C. The crystals were transferred into reservoir solution supplemented with 20% (vol/vol) glycerol for cryoprotection and flash-frozen in liquid nitrogen.

Data Collection, Structure Determination, and Refinement of the Aae ΔNeN/DNA Complex. The crystals belonged to space group P3₂₁ (SI Appendix, Table S1), but analysis using XTRIAGE within PHENIX (31) revealed that the crystals were often twinned (operator h, −k, −l), with variable twinning fractions. A large number of crystals were screened to find the best diffracting datasets that were also not twinned. Phasing information was obtained using TauBr14 (57).
Rebuilding and Re-Refinement of $\sigma^E$ Using the Deposited Structure Factors Associated with PDB 5BYH. As a starting model, we used coordinates of Eco core RNAP derived from Eco $\Delta \sigma E$-holoenzyme with $\Delta \sigma E$ removed (PDB ID code 4LJZ) (58) superimposed onto the Eco core RNAP of PDB 5BYH (13). For initial models of the $\kappa_{EM}$ $\sigma^E$ domains (CBD, $\Delta \sigma E$, and $\Delta \sigma E$-protonated) in the cell, we used homology models using the $\sigma^E$ $\sigma^E$ domains as templates using the Swiss Model Server (59).

ACKNOWLEDGMENTS. We thank M. Edghill, D. Nixon, L. V. Kelley, and M. de la Maza, and D. Baker for initial contributions to this project, and B. T. Nixon for helpful discussions. This work was based in part, on research conducted at the National Synchrotron Light Source. This work was supported, in part, on research conducted at the National Synchrotron Light Source. This work was supported by a Women in Science Fellowship at The Rockefeller University. This work was supported by NIH Grant R35 GM181130 to (S.A.D.).